

The pursuit of computational justice in open systems

Jeremy Pitt · Dídac Busquets · Régis Riveret

Received: 15 March 2013 / Accepted: 28 November 2013 / Published online: 24 December 2013
© Springer-Verlag London 2013

Abstract Many open networks, distributed computing systems, and infrastructure management systems face a common problem: how to distribute a collectivised set of resources amongst a set of autonomous agents of heterogeneous provenance. One approach is for the agents themselves to self-organise the allocation of resources with respect to a set of agreed conventional rules; but given an allocation scheme which maps resources to those agents and a set of rules for determining that allocation scheme, some natural questions arise—Is this allocation fair? Is the allocation method effective? Is it efficient? Are the decision makers accountable? In this paper, we argue that some answers to these questions can be found in the formal characterisation of different aspects of ‘justice’ and that these different aspects need a principled operationalisation as policies for system management. We present a formal model and some experimental results, concluding that the different aspects are all inter-connected and that what is required is a comprehensive research programme in computational justice.

Keywords Multi-agent systems · Self-organisation · Resource allocation · Computational justice

1 Introduction

In open systems, it is a common requirement to collectivise and distribute computing resources amongst the components of the system. This requirement cuts across different scales of time and space and is a feature of distributed systems for cloud and grid computing (Ardagna et al. 2011; Birman et al. 2009; Choi et al. 2008), and sensor and vehicular networks (Ding et al. 2003; Manvi et al. 2009). The decision-making required to achieve this distribution is too fast, too frequent and too complex for (human) operator intervention, so the system components have to self-organise the distribution by, and between, themselves.

Furthermore, there is another trend towards the automation of infrastructure management systems, for example in energy grids, water irrigation and transportation systems, which critically involve active participation of (human) users (Bourazeri et al. 2012; Ferscha et al. 2011). In these applications, there is also a requirement to share physical resources amongst the infrastructure users, who can be both producers and consumers of resources (i.e. prosumers, typically found in energy grids). If that infrastructure is instrumented with inter-connected sensors and (personal) devices, producing a so-called smart infrastructure, then the (user-centric) self-determination of the resource distribution can be assisted by computational means.

In both cases, whether a ‘technical’ system or network composed purely of autonomous computing components, as found in grid computing or sensor networks, or a socio-technical system composed of ICT-enabled people interacting with inter-connected, instrumented (and increasingly intelligent) devices, such as Smart Grids, we see that the actors have to self-determine the resource allocation; however, they can also self-determine the rules that are used in this self-determination. This is self-organisation: a

J. Pitt (✉) · D. Busquets · R. Riveret
Department of Electrical and Electronic Engineering, Imperial
College London, London SW7 2BT, UK
e-mail: j.pitt@imperial.ac.uk

D. Busquets
e-mail: didac.busquets@imperial.ac.uk

R. Riveret
e-mail: r.riveret@imperial.ac.uk

group of interacting entities who come together for some collective purpose and agree amongst and by themselves, who is (and is not) a member of the group; what rules should be followed to achieve the collective purpose; what rules should be used to change the rules intended to achieve the collective purpose, and so on.

Given a set of actors needing to share resources, an allocation scheme which maps resources to those actors, and a set of rules for determining that allocation scheme, some natural questions arise—Is this allocation fair? Is the allocation method effective? Is it efficient? Are the decision makers accountable? To what extent did those affected by the rules participate in their selection? Was any punishment for non-compliance with the rules proportional to the severity of the offence?

In this paper, we argue that some answers to these questions can be found in the formal characterisation of different aspects of ‘justice’ and that these different aspects need a principled operationalisation as policies for system management. We present a formal model and some experimental results and conclude that the different aspects are all inter-connected and that what is required is a comprehensive research programme in computational justice.

Accordingly, this paper is structured as follows. In Sect. 2, we identify key features of prototypical applications of self-organising systems for ‘technical’ or socio-technical applications. This motivates the analysis of multiple different strands of justice discussed in Sect. 3. These strands of justice are woven together in Sect. 4, using a common formal characterisation in the framework of self-organising electronic institutions (Pitt et al. 2012), an approach to self-determined resource allocation, role assignment and rule selection/modification grounded on the principles of Ostrom’s self-governing institutions (Ostrom 1990). Section 5 presents a summary of several experiments with the principled operationalisation of ‘justice’ as executable ‘policies’ and expose the tangled nature of the strands of justice in relation to Ostrom’s principles. Finally, in Sect. 6, we outline a broader research programme of *computational justice* and expose a set of deeper issues for computational justice in open self-organising systems to be addressed by future research.

2 Self-organising open systems

We characterise *open systems* in the sense of Hewitt (1986): there is a set of autonomous agents of heterogeneous provenance; it can be assumed that they are able to communicate via a common language, but it cannot be assumed that there is a centralised controller that the agents are cooperative and share common goals or that any agent will comply with the system specification.

In this section, we survey some open ‘technical’ and socio-technical systems in which the self-organisation of resource allocation, role assignment and/or rule selection/modification is a crucial aspect of their successful operation. From this survey, we extract a set of key common features.

2.1 Prototypical applications: ‘technical’ systems

2.1.1 Grid computing

The idea of grid computing is to provide computing services ‘on demand’, and in particular the Desktop Grid provides a distributed computing infrastructure based on communal pooling of resources (Choi et al. 2008). This enables agents (on behalf of their users) to offer or exploit idle computing resources on (respectively) their own or other machines to process large, long-running, computationally intensive tasks in parallel, thereby increasing throughput, not wasting CPU cycles, and so on. In the Trusted Desktop Grid, the problem of free-riding (i.e. exploiting pooled resources without contributing to the pool) was addressed by the self-organisation of a ‘trusted community’ (Bernard et al. 2011).

2.1.2 Cloud computing

One application of multi-tenant cloud computing is to support real-time on-demand provisioning of different types of computing facility, software, platform, infrastructure, etc., undifferentiated as services, hence software-as-a-service, platform-as-a-service, infrastructure-as-a-service, and so on (Birman et al. 2009). To reduce the total cost of ownership to the clients and to maximise revenue, the cloud service provider requires as many clients to use the cloud offering, without overloading it such that quality-of-service would be diminished and contractual penalties incurred. In one sense, this is a generalisation of well-studied optimisation problems such as job-shop scheduling, except that the ‘jobs’ and the ‘shops’ are autonomous decision makers and availability of both may change over time. Self-organising both tenancy arrangements and the ‘schedule’ (i.e. the resource allocation) is one proposed approach to resolving this tension (Pitt et al. 2011).

2.1.3 Sensor networks

A sensor network is a type of open system with resource constraints, in which functionality (i.e. distributed data measurement, aggregation and reporting) may be compromised by a lack of resources, specifically battery power. In particular, a trade-off exists between accuracy and longevity for networks implementing in-network data

aggregation functions. Given that the major drain on battery power is in communication, one solution is to self-organise the network into clusters to minimise routing and transmission overheads; then self-organise the clusters to distribute power consumption whilst staying within an upper bound for the cluster reporting error. The idea is to achieve a ‘synchronised failure’ with reliable reporting until the time of network failure with the least power left unused, rather than a prolonged degradation with increasing unreliability and for the network to fail when some nodes still have power.

2.2 Prototypical applications: socio-technical systems

2.2.1 Smartgrids

The vision for the energy grid of Schönau was a decentralised form of green-energy production, in terms of both increasing the efficiency of energy transmission and empowering citizens to take charge of their energy consumption and production (<http://www.ews-schoenau.de>). The initial idea was to turn energy consumers into prosumers, a combination of producers and consumers, by motivating individuals to produce and save energy, and to sell the surplus back to the grid. This way of thinking initiated the process of equipping the inhabitants of Schönau with resources to produce energy and manage it through a citizen-owned social business, Power Supplier of Schönau. Most households in this community produced energy by diverse means, and managed the process of its distribution. However, to achieve this, a change in the law was demanded by the community so that the inhabitants could become the owners of their part of the energy grid and the managers of the distribution process.

2.2.2 Giffgaff

Giffgaff is a UK-based mobile phone service which can be differentiated from other operators’ services because giffgaff subscribers can also participate in specific aspects of the network operation, in particular sales and marketing, customer services and product testing. For participation in these activities, subscribers earn remuneration in the form of ‘Payback’, which are points that can be redeemed for cash, credit or charitable donations.

2.2.3 Wikipedia

Many other social networking and user-generated content-management platforms demonstrate a similar pattern of user engagement and active participation, sometimes in return for a digital currency, sometimes for social capital. However, the Wikipedia case is unique: user types and

associated roles (privileges and responsibilities) are not preset by either the platform itself or by the owner of the service. Instead, the rules governing who is responsible for what resource and empowered to perform which actions are determined by the users themselves. Anybody can propose a new rule or a rule change—including the rules how to choose moderators and their privileges and how to change rules. The community decides whether to adopt the proposal or not. Wikipedia is an evolving, self-organising institution adapting to the needs of its users: it is also likely that some mobile network services and social networking platforms will follow suit.

2.3 Key features

All the prototypical application considered above are *open*, that is to say, that they are composed of autonomous entities of heterogenous provenance; there is no central controller; there is not necessarily a common goal (as distinct from a common purpose), so the entities may be competing, in particular for resources; but both a common ‘language’ and a mutually agreed and understood rule-set specifying ‘socially acceptable’ behaviour can be assumed. Entities (agents) can join and leave the system, in particular some entities may leave for good and some trying to join may never have been previously encountered.

However, in addition to openness, these prototypical applications also exhibit a common set of key features, which we identify as:

- *self-determination* in the absence of a central controller, resource allocation, the rules for determining the resource allocation, etc., must be determined by the entities themselves;
- the *expectation of error* in the presence of competition and conventional rules, sub-ideal behaviour (contrary to specification) is to be expected, but errors may be a result of accident or necessity as well as malice;
- *enforcement* these systems might as well use random allocation if agents can repudiate conventional rules and sanctions for non-compliance by refusing to abide by their outcomes;
- an *economy of scarcity* (cf. Rescher 1966) there are sufficient resources to keep the appropriators ‘satisfied’ in the long term, but insufficient resources to meet everyone’s demands at any a particular time-point;
- *endogenous resources* in a system where all the resources are provided by the appropriators themselves, as in a sensor network or a micro-grid, computing the resource allocation must be ‘paid for’ from these resources; and
- *no full disclosure* the appropriators are autonomous and internal states cannot be checked for compliance (with

conventional rules), so incoming agents do not have all the information required for necessarily reliable investment decisions (e.g. contributing to a common pool).

It is our contention that, if we are dealing with ‘intelligent’ entities capable of representing and reasoning with conventional rule-sets, then these features can be addressed by different aspects of ‘justice’.

3 Justice (five different qualifiers)

There are many different qualifiers of the term ‘justice’. In this section, the five we consider in turn are natural, distributive, retributive, procedural and interactional justice. We then relate these five qualifiers of justice to the key features of open systems identified in the previous section.

3.1 Natural justice

The term natural justice in UK law is derived from Roman principles of justice which were assumed to be self-evident or axiomatic and did not need a statutory representation. Such principles were *audi alteram partem* (hear the other side) and *nemo iudex in parte sua* (no-one a judge in their own case). More recently, in his book of the same title, Binmore (2005) uses evolutionary game theory to investigate how groups of individuals can come to social arrangements that are fair, equitable and stable, as part of a broader ambition of establishing a science of morality based on the theory of games. However, a third definition of natural justice comes from the Human Rights principle of Participation and Inclusion that “All people have the right to participate in and access information relating to the decision-making processes that affect their lives and well-being”, which is a founding principle of the international non-profit organisation Natural Justice (<http://naturaljustice.org>).

This third definition is strongly correlated with Elinor Ostrom’s principles for enduring common-pool resource management using self-governing institutions. It had been claimed that people sharing access to a common resource will inevitably act in such a way as to deplete the resource in the short term, even if it is no-one’s interest in the long term, the so-called tragedy of the commons (Hardin 1968). Ostrom’s thesis (for which she was awarded the Nobel Prize for Economic Science in 2009) was that in managing a common-pool resource, the tragedy of the commons was not inevitable, and indeed was eminently avoidable. Through extensive fieldwork and case studies, she showed that on numerous occasions groups of people managed collectively to avoid the tragedy by forming an institution,

Table 1 Ostrom’s principles for enduring institutions

1	Clearly defined boundaries
2	Congruence between appropriation and provision rules and the state of the prevailing local environment
3	Collective-choice arrangements: those affected by the provision and appropriation rules participate in selection and modification of those rules
4	Monitoring by accountable agencies
5	Graduated sanctions
6	Access to fast, cheap conflict-resolution mechanisms
7	Existence of and control over their own institutions is not challenged by external authorities
8	Systems of systems

which specified conventional rules that regulate and coordinate people’s behaviour.

However, her fieldwork also showed that just forming an institution was not enough, there were cases when the institution ensured that the resource was sustained, and other cases where it did not. From this she identified a set of eight institutional design principles which were essential and determinate conditions for a self-governing institution; a subsequent meta-review has corroborated these principles with only minor clarifications (Cox et al. 2010). These conditions are summarised in Table 1. The principles most pertinent to the third definition (above) of natural justice are Principle 1, clearly defined membership; and Principle 3 that those affected by institutional rules regulating provision to, and appropriation from, a common-pool resource should participate in the specification, selection and modification of those rules.

3.2 Distributive justice

Distributive justice dates back to Ancient Greece, with Aristotle’s maxim, also referred to as principle of distributive justice, stating that “equals should be treated equally, and unequals unequally, in proportion to the relevant similarities and differences”. Since then, many different theories about the subject have been developed. These can be classified into three main families: *equality and need*; *utilitarianism and welfare economics*; and *equity and desert*.

The first family, *equality and need*, is characterised by its concern for the welfare of those least advantaged in the society. This inspires the *need principle*, which seeks for an equal satisfaction of basic needs. Theories in this family include egalitarianism, Rawls’ theory of justice (1971), and Marxism, among others.

The second family, *utilitarianism and welfare economics*, relies on the *efficiency principle*, which seeks maximising the global surplus (also referred to as outcome,

utility, satisfaction) of the society. Hence, it does not deal with individual outcomes, but in the aggregation of these. In this family we can find, among others, the utilitarianism theory (Bentham 1789), Pareto principles theories and envy-freeness.

The third family, *equity and desert*, advocates for a dependence of allocations on the actions of each individual, according to the *equity principle*. This principle states that an individual should receive an allocation that is proportional to her contributions (either positive or negative) to the society. Theories in this family include equity theory, desert theory and Nozick's theory (1974).

A more detailed description of these families and the different theories in each of them can be found in (Konow 2003). In this work, Konow concludes that the choice of a justice theory should not be limited to one belonging to one family or another, but that different theories may be combined, in what he calls *pluralistic justice*.

This coincides with Rescher's (1966) analysis on distributive justice, who concluded that the Principle of Utility, taken as a fairness metric expressed as "the greater good of the greater number", is but one of many prevailing considerations which need to be taken into account when determining a 'fair' allocation of resources. Rescher (1966) then observed that distributive justice had been held, by various sources, to consist of treating people wholly or primarily according to one of seven *canons* (established principles expressed in English), as the canons of equality, need, ability, effort, supply and demand, productivity and social utility.

Rescher's analysis showed that each canon, taken in isolation, was inadequate as the sole dispensary of distributive justice. Instead, his position was that distributive justice was found in the *canon of claims*, which consists of treating people according to their legitimate claims, both positive and negative. As Rescher claimed, this re-directed the search for distributive justice towards determining what the legitimate claims are, how they are accommodated in case of plurality, and how they are reconciled in case of conflict.

Lately, some work developed in Social Sciences has been successfully transposed to multi-agent systems via Computational Social Choice for distributed resource allocation and negotiation (Chevalere et al. 2007). A specific concern is to define metrics and design computational models that encourage (or compel) rational agents to determine an optimal or fair allocation of resources (i.e. mechanism design).

3.3 Retributive justice

There is a cultural tendency according to which, when an agent does something wrong, such as non-compliance with

a normative system, then the wrongdoer should deserve punishment and, on the contrary, good behaviour should be rewarded.

The philosophers of punishments traditionally distinguish retributivism and utilitarianism. The retributivist perspective holds that punishment is a necessary consequence of committing an offence (the punishment of past wrongdoing). Retributivists assert that non-moral agency deserves to be punished in order to maintain a moral order. For example, Kant asserted that no principle but retribution was a legitimate basis for punishment. This contrasts with the utilitarian notions of justice according to which punishment is justified by its potential benefits and in particular the avoidance of future wrongdoing. From the most common utilitarian perspective, a punishment should bring some loss of utility to the wrongdoing agent; and the justification for the punishment is that this loss of utility acts as a deterrent to future wrongdoing, both to prevent recidivism (the same agent repeating the wrongdoing) and to encourage other agents to refrain from offending.

Whatever justification for punishment is established, practical questions then arise regarding the design of a punishment system. This includes the forms of punishment, e.g. whether the sanction should be a fine or the suspension of a licence, and the principles of punishment, e.g. a commonly held principle is that a punishment should be proportionate to the severity of the wrongdoing. In this view, Ostrom (1990) maintained that a system of *graduated sanctions* was a necessary condition for enduring commons, but also that there should be a fast and effective dispute-resolution system. There are also different compensation strategies for *restorative justice*, i.e. making good for loss suffered as a consequence of wrongdoing and to favour forgiveness. As a concrete example of possible restorative alternatives, when there is a breach of contract, one may consider the position where the victim would have been in if the contract had been fulfilled or, alternatively, if the contract had never been signed.

3.4 Procedural justice

Robert's Rules of Order (Robert et al. 2000) is a comprehensive manual of procedures for conducting business in a deliberative assembly, i.e. a group of individuals making a decision about some policy or course of action, and mutually agreeing some conventional rules and procedures which regulate how that decision is to be made. Procedural justice is concerned with ensuring that those rules and procedures are fit-for-purpose, and includes elements of 'fairness', 'transparency' and cost.

Concepts of procedural justice, i.e. what makes a procedure fair, are of concern to many other fields of human endeavour as well as political deliberation, including

dispute resolution in law, public health, organisational psychology, and philosophy. This has led to a number of different definitions, for example, procedural justice is ...:

- ... a theory of procedural fairness for civil dispute resolution (Solum 2004);
- ... a requirement for a community to engage in a democratic process to determine which public health functions the authorities should maintain, with respect to a trade-off between costs and benefits (Kass 2001) and a justification for their decisions (Uphsur 2002);
- ... a four-component model derived from the interaction of procedural function and source (Blader and Tyler 2003);
- ... subject to a graded distinction between perfect, imperfect and pure procedural justice (Rawls 1971).

Solum (2004) contends that a theory of procedural justice must address two problems, an ‘easy’ and a ‘hard’ one. The ‘easy’ problem is to produce accurate outcomes at acceptable cost, what has been referred to as the balancing model of procedural justice. The ‘hard’ problem is concerned with solving how, if perfect accuracy is unattainable, people feel compelled to comply with what they perceive as a mistaken judgement. The resolution to this issue is for two principles of procedural justice to work in harmony: the *participation* principle, which ensures that the procedural arrangements provide ‘adequate’ participation, and the *accuracy* principle, which ensures that the procedural arrangements provide ‘maximum’ likelihood of achieving the ‘correct’ outcome.

Kass (2001) argues that achieving population-oriented (macro-level) goals in public health (as opposed to individual goals) requires a trade-off between the benefits and effectiveness of a proposed health policy to the population as a whole against the burdens (costs) and possible infringements of civil liberties imposed on the individuals separately. Fair procedures should be used to determine which burdens are acceptable to the community in return for the benefits.

Also in the field of public health, Uphsur (2002) cites the *transparency* principle, that public health authorities are obliged to communicate the grounds for their actions and decisions, and allow an appeals process.

The four-component model of procedural justice (Blader and Tyler 2003) contends that *people* use four types of judgement to evaluate group procedures. These judgements involve evaluating how the formal rules (for group procedures) treat group members and how decisions are made; and evaluating how the group authorities make decisions and treat group members. As such this is a relational model of procedural justice based on subjective assessments of procedural function (decision-making) and its source (the group authorities).

Rawls (1971) differentiates between *perfect* procedural justice, which holds if there is an independent criterion for determining a fair outcome of a procedure and a method which guarantees that the fair outcome will result from execution of the procedure. *Imperfect* procedural justice exists if there is a such a criterion, but no method that is guaranteed to produce it; and *pure* procedural justice exists if there no criterion but the procedural method itself.

3.5 Interactional justice

Interactional justice comprises two specific types of interpersonal treatment (Greenberg 1993). The first is interpersonal justice, which is a measure of how ‘well’ those who are responsible for executing procedures or determining outcomes treat those who are subject to the procedures and outcomes. The second is informational justice, which is a measure of the justifications provided to the subjects about which procedures were used or why resources were allocated as they were.

3.6 Justice and the key features

We are now in a position to state how different qualifiers of justice can be used to address the key features of open self-organising systems previously identified:

- self-determination requires a concept of natural justice, namely the right of participation and inclusion, usually by voting;
- recognising the occurrence of errors and the enforcement of sanctions (and agreements reached by conventional rules) requires a concept of retributive justice: distinguishing between different types of error and offering the chance of redemption and allowing for appeals are also essential;
- in an economy of scarcity, familiar fairness and efficiency criteria, like Pareto efficiency and envy-freeness, may be ineffective in the long term, and a concept of distributive justice (or outcome justice) and a subjective agreement on fairness norms is required (Elster 1992);
- dealing with endogenous resources requires a concept of procedural justice: if the administration of the rules has to be ‘paid for’ from the same resources that are otherwise allocated for ‘useful’ jobs, then it is necessary to ensure that they are efficient; and
- dealing with lack of full disclosure requires an element of interactional justice, namely informational justice, to force disclosure of relevant information.

The next step is to weave the five different strands of justice together and represent them in computational form: our proposal for achieving this is through *self-organising electronic institutions*.

4 The formal characterisation of ‘justice’

In this section, we review Ostrom’s definition of an institution, and introduce a computational framework for specifying norm-governed systems. We define a self-organising electronic institution as a formal model for Ostrom’s definition, and propose to give a computational representation of the institutional rules using an action language (specifically the Event Calculus (EC) Kowalski and Sergot 1986). This will provide the basis for policy specifications of the different qualifiers of justice, as described in the following section.

4.1 Self-governing institutions

Ostrom (1990) defined an institution as a set of working rules that specified procedures for operational-, collective- and constitutional-choice procedures. These procedures were, respectively, concerned with:

- *operational choice* rules for provision, allocation and appropriation; monitoring; and access control;
- *collective choice* rules for specifying, selecting and adapting the operational-choice rules; rule enforcement and dispute resolution; and
- *constitutional choice* rules for specifying the eligibility (of institution members) for determining the collective-choice rules.

These rules were role-based, mutually agreed, mutable and nested within each other in *action situations*. Distinguishing between nested action situations requires a formal characterisation of institutionalised power (Jones and Sergot 1996), whereby a designated agent appointed to an identified role in an action situation is empowered to bring about a fact of conventional (or institutional) significance by performing a recognised action in that specific situation.

These rules can be given a formal characterisation simply as ordinary functions, mapping specific inputs onto specific outputs (see below for examples).

4.2 Dynamic norm-governed multi-agent systems

The study of legal, social and organisational systems has often been formalised in terms of norm-governed systems. Artikis (2012) introduced a framework that allows agents to modify the rules or protocols of a norm-governed system at runtime. This framework defines three components: a specification of a norm-governed system; a protocol stack for defining how to change the specification; and a topological space for expressing the ‘distance’ between one specification instance and another.

In more detail, the specification of a norm-governed system expressed five aspects of social constraint: the

physical capabilities; the institutionalised powers; the permissions, prohibitions and obligations of the agents; the sanctions and enforcement policies that deal with the performance of prohibited actions and non-compliance with obligations; and the designated roles of empowered agents.

The second component of the framework is a communication language which is used to define a set of protocols for conducting the business of the institution. In the framework, the protocol stack is used by the agents to modify the rules or protocols of a norm-governed system at runtime. This stack defines a set of object-level protocols, and assumes that during the execution of an object protocol the participants could start a meta-protocol to (try to) modify the object-level protocol. The participants of the meta-protocol could initiate a meta-meta protocol to modify the rules of the meta-protocol, and so on. In addition to object- and meta-protocols, there are also ‘transition’ protocols. These protocols define the conditions in which an agent may initiate a meta-protocol, who occupies which role in the meta-protocol, and what elements (the degrees of freedom (DoF)) of an object protocol can be modified as a result of the meta-protocol execution.

Each modifiable parameter of a protocol is a DoF which can take one value from a set of values. Taking one value for each DoF defines an *instance* of the specification. All the possible combinations of different values for each DoF define a set of instances. This set can be turned into a metric space by defining a distance metric d which can determine how different (the ‘distance’) between one specification instance and another. It is also possible to define social constraints about moving in this space, e.g. that some specification instances are prohibited that a change from one specification instance to another cannot exceed a certain distance, and so on.

4.3 Self-organising electronic institutions

The framework of dynamic norm-governed systems can be used to specify a wide variety of such systems; however, we are only interested in a subset of these systems. Accordingly, we define a Self-Organising Electronic Institution as a collection of agents plus a specification of a dynamic norm-governed system which encapsulate a set of operational-, collective and constitutional-choice rules, and the associated action situations, for realising self-organisation, self-regulation, and other self-* properties (i.e. the electronic ‘equivalent’ of Ostrom’s self-governing institutions).

Formally it can be denoted by \mathcal{IC}_t , which is a multi-agent system at time t defined by:

$$\mathcal{IC}_t = \langle \mathcal{A}, \mathcal{C}, \mathcal{R}, \mathcal{L} \rangle_t$$

where (omitting the subscript t if clear from context):

- \mathcal{A} is the set of all agents;
- \mathcal{C} is the set of action situations;
- R is a binary *nesting* relation on \mathcal{C} ;
- \mathcal{L} is a dynamic norm-governed system specification.

In the framework of (Artikis 2012), a number of degrees of freedom (DoF) are identified, so \mathcal{L} defines a specification space, where each specification instance is defined by a different set of values assigned to the DoF. For example, one degree of freedom is the operational-choice rule by which the resources are allocated (e.g. at random, by ration, or using legitimate claims); another degree of freedom is the winner-determination method used to elect an agent to occupy a role; another degree of freedom is the appointment to the role itself. Some of the degrees of freedom may be changed by rules contained within the action situation, and some may be changed by decisions made in the action situation with which it is nested, as determined by the nesting relation R .

Each action situation $C_{i,t} \in \mathcal{C}_t$ is defined by:

$$C_{i,t} = \langle \mathcal{M}, l, \epsilon \rangle_t$$

where (again omitting the subscripts as clear from context):

- \mathcal{M} is the set of members such that $\mathcal{M} \subseteq \mathcal{A}$
- l is a specification instance of \mathcal{L} ; and
- ϵ is the cluster's local environment, a pair $\langle Bf, If \rangle$.

Regarding the environment ϵ , Bf represents the set of 'brute' facts whose values are determined by the *physical* state, including the sum of common-pool resources P as a result of provision by the agents. If represents the set of 'institutional' facts, whose values are determined by the *conventional* state, i.e. are asserted by the exercise of institutionalised power.

4.4 Computational representation of institutional rules

In this framework, the rules of \mathcal{L} can be used to give a principled operationalisation of the qualifiers of justice as policies for system management.

One way to do this is to represent the rules of \mathcal{L} in the EC (Kowalski and Sergot 1986). The EC is a logical formalism for representing and reasoning about actions or events and their effects based on a many-sorted first-order predicate calculus. An *action description* in EC includes axioms that define a narrative (the occurrence of actions), using the *happensAt* predicate; the effects of actions, using *initiates* and *terminates* predicates; and the values of the fluents, using *initially* and *holdsAt* predicates. A

fluent is then a proposition whose values can change over time.

\mathcal{L} is then an EC action description containing axioms of the form:

$$\text{Action initiates } F = V \text{ at } T \leftarrow \text{Conditions}$$

which are read as stating that the occurrence of action *Action* at time T initiates a period of time for which the value of fluent F is V , if the *Conditions* are satisfied. The physical and institutional facts in an action situation's environment ϵ are represented as EC fluents.

Generally, the conditions will include that the agent performing the action is empowered (has the institutionalised power) to do so. This is represented as:

$$\text{pow}(\text{Agent}, \text{Action}(\dots)) = \text{true holdsAt } T$$

Note that (institutionalised) power to perform an action often, but not always, implies permission to perform that action.

A *narrative* is a sequence of actions which happen at specific times:

$$\text{Action}_1 \text{ happensAt } T1$$

$$\text{Action}_2 \text{ happensAt } T2$$

...

As an executable specification, given an EC action description, a narrative of events and a starting state at time t of an action situation $C_{i,t}$, using the EC engine the specification can be queried to determine what was the state of the action situation at some later time-point t' , $C_{i,t'}$ (i.e. what *holdsAt* when), and indeed at every time-point in-between. In this way, the EC can be used as both a specification of a policy and to check that the execution (carrying out) of the policy has conformed to its prescriptions (i.e. the specification can be animated to validate the 'correctness' of the specification).

In the next section, we show how policy specifications can be representing using \mathcal{L} (i.e. the EC), and these policies can be used to capture different qualifiers of justice.

5 Experiments with self-organising electronic institutions

In this section, we analyse separate experiments with qualifiers of justice in self-organising electronic institutions. The first experiment is concerned with Principles 1 and 3 about clearly defined boundaries for membership, and the policy for natural justice is encoded in role-assignment protocols and voting protocols (Pitt et al.

2011). The second experiment describes a preliminary investigation of self-organisation of a system of graduated sanctions against the frequency of offences, as a test of a self-organised system of retributive justice, interleaving Principle 3 with Principles 4/5/6 (Pitt and Schaumeier 2012). Finally, the third experiment reports on an encoding of Rescher’s theory of legitimate claims in the context of self-organised resource allocation in an economy of scarcity, as a policy for distributive justice, and concerns Principles 2 and 3 (Pitt et al. 2012).

5.1 Natural justice

5.1.1 Experimental setting

This experiment with natural justice concerned the appropriation and provision of endogenously generated resources and multiple institutions, in the context of collective-choice arrangements (Principle 3) for institutional membership (Principle 1).

This problem has typically been analysed as a linear public good (LPG) game (Gaechter 2006). In a typical LPG game, n people or agents form a group or *cluster*. All cluster members individually possess a quantity of resource. Each cluster member $i \in \{1, \dots, n\}$ decides independently to contribute resources $r_i \in [0, 1]$ to the public good. The contributions from the whole cluster are summed and the payoff u_i for each player i is given by:

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

The first term represents the payoff from the public good (the ‘public payoff’), distributed equally among the n cluster members. The second term represents the payoff from the resources withheld from the public good (the ‘private payoff’) irrespective of how much was contributed individually and collectively. The coefficients a and b represent the relative value of the public/private payoffs, respectively. Note that if the conditions on a and b hold, a rational but selfish agent has the incentive to contribute 0 to the public good.

In each cluster, this game is ‘played’ in iterated rounds with one rule for compliance: that an agent must not provision fewer resources than the average for the cluster of which it is a member.

To play this game in the context of an institution, the formal characterisation of the institutional rules includes two operational-choice rules (ocr_i) for role assignment,

(with, in parenthesis, the role responsible for its enactment and enforcement), and four collective-choice rules (scr_i) for assigning an agent to a role, and for choosing an access control method ($acMethod$) and exclusion method ($exMethod$):

$$\begin{aligned} (gatekeeper)ocr_1 &: \mathcal{M}^c \times acMethod \rightarrow Bool \\ (monitor)ocr_2 &: \mathcal{M} \times V(\cdot)_{a \in \mathcal{M}} \times exMethod \rightarrow Bool \\ (head)scr_1 &: V(\cdot)_{a \in \mathcal{M}} \times wdMethod \rightarrow \mathcal{M} \\ (head)scr_2 &: V(\cdot)_{a \in \mathcal{M}} \times wdMethod \rightarrow acMethod \\ (head)scr_3 &: V(\cdot)_{a \in \mathcal{M}} \times wdMethod \rightarrow \mathcal{M} \\ (head)scr_4 &: V(\cdot)_{a \in \mathcal{M}} \times wdMethod \rightarrow exMethod \end{aligned}$$

where $V(\cdot)_{a \in \mathcal{M}}$ is a set of expressed preferences on an issue by each *member* agent in cluster I , where $I \in \mathcal{I}$, and so on.

Thus, the collective-choice rules scr_1 and scr_2 use a winner-determination method ($wdMethod$) to map a set of votes onto, respectively, an agent (who is appointed to the role of *gatekeeper*) and one of the two access control methods. Similarly, the social collective-choice rules scr_3 and scr_4 use a winner-determination method to map a set of votes onto, respectively, an agent (who is appointed to the role of *monitor*) and one of the two exclusion methods. These collective-choice arrangements are all applied by the agent occupying the *head*.

The operational-choice rule ocr_1 is applied by the *gatekeeper* to map an application to join from an agent not in \mathcal{M} (i.e. the set complement \mathcal{M}^c) to a boolean outcome depending on the access control method. A true result means the applicant can be assigned the role of *member*. Similarly, the rule ocr_2 is applied by the *monitor* to map an agent in \mathcal{M} that did not comply with the rules of the LPG game to a boolean outcome using the exclusion method.

5.1.2 Specification space

These rules in \mathcal{L} effectively define four DoF: the selection of the $acMethod$ and the selection of the $exMethod$, and the assignment to the *gatekeeper* role and the assignment to the *monitor* role.

Since no agent can occupy both roles *monitor* and *gatekeeper* in a cluster I with n members, it follows that there are $4n^2 - 4n$ possible specification instances. Rather than dynamically computing the entire space for each cluster and trying to determine the ‘optimal’ configuration, we separate the ‘specification instance’ selection function into two dimensions.

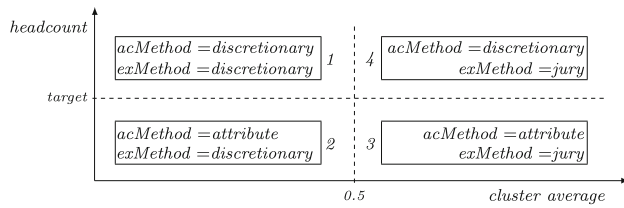


Fig. 1 Specification space

For the first dimension, the decision of which agent to assign to the role of *monitor* or *gatekeeper*, we define a family of preference functions, some based on relevant properties of the agent (e.g. compliance probability, time already spent in the role, etc.) and some not (e.g. random, nominative proximity, etc.). Each agent is associated with a subset of these functions, and applies them when voting for either *monitor* or *gatekeeper*. The *head* is empowered to assign the role to the agent with the most votes according to a winner-determination method.

For the second dimension, the selection of *acMethod* and *exMethod*, we define two criteria. The first criterion is a *target* membership: this value is a trade-off between total cost of ownership (which is too high if the *headcount* is less than the target) and the quality-of-service (which is too low if the *headcount* is more than the target). The second criterion is the average probability of compliance in the LPG game. For each agent, we define a probability distribution for voting for a change in the specification according to these criteria. As a result, the selected specification instance falls into one of the quadrants 1–4 as shown in Fig. 1.

We note, in passing, that the evaluation of the ‘performance’ of the cluster with respect to the indicated criteria can be considered as a rudimentary form of procedural justice, being used throughout the collective-choice rules to configure the operational-choice rules of membership (which is the system of natural justice). However, ‘performance’ is a multi-faceted measure encompassing many different kinds of (weighted) indicators, and in the context of organisations appropriate metrics turn out to be highly problematic both to specify, measure and interpret.

5.1.3 Policy specification

The EC specification of the operational-choice rule ocr_1 is given by a role-assignment protocol for membership. An agent can apply for membership to a cluster I if it does not occupy a role in any other cluster:

$apply(A, I)$ initiates $applied(A, I) = true$ at $T \leftarrow$
 $not\ role_of(A, _) = member$ holdsAt T

The gatekeeper agent is empowered to admit the agent, to the cluster, by an *assign* action, depending on the access control method.

$assign(G, A, member, I)$ initiates

$role_of(A, I) = member$ at $T \leftarrow$

$pow(G, assign(G, A, member, I)) = true$ holdsAt T

$pow(G, assign(G, A, member, I)) = true$ holdsAt $T \leftarrow$

$applied(A, I) = true$ holdsAt $T \wedge$

$acMethod(I) = attribute$ holdsAt $T \wedge$

$role_of(G, I) = gatekeeper$ holdsAt $T \wedge$

$role_conditions(member, A, I) = true$ holdsAt T

$pow(G, assign(G, A, member, I)) = true$ holdsAt $T \leftarrow$

$applied(A, I) = true$ holdsAt $T \wedge$

$acMethod(I) = discretionary$ holdsAt $T \wedge$

$role_of(G, I) = gatekeeper$ holdsAt T

For the game rule, this can be defined by a rule on the provide action, assuming this can be monitored and monitoring is for free (but see below).

$provide(A, R, I)$ initiates $provision(A, I) = R$ at $T \leftarrow$

$pow(A, provide(A, R, I)) = true$ holdsAt T

$provide(A, R, I)$ initiates

$sanctioned(A, T, I) = true$ at $T \leftarrow$

$pow(A, provide(A, R, I)) = true$ holdsAt $T \wedge$

$provision(A, I) = R$ holdsAt $T \wedge$

$cluster_average(I) = Ave$ holdsAt $T \wedge R < Ave$

$pow(A, provide(A, R, I)) = true$ holdsAt $T \leftarrow$

$role_of(A, I) = member$ holdsAt T

The *monitor* is empowered to exclude a *member* that does not comply with the rules of the game G . For each iteration of G , agents should contribute resources in the interval $[ave_I, 1]$ to comply, where ave_I is the average contribution of resources from the previous iteration [the value of the fluent $cluster_average(I)$].

If the *exMethod* is discretionary, then the *monitor* agent G can exclude the applicant A (or not) as it decides. If the *exMethod* is jury, then the *monitor* must have called for a vote on the issue of the exclusion of the applicant:

$exclude(G, A, member, I)$ initiates
 $role_of(A, I) = none$ at $T \leftarrow$
 $pow(G, exclude(G, A, member, I)) = true$ holdsAt T
 $pow(G, exclude(G, A, member, I)) = true$
holdsAt $T \leftarrow$
 $role_of(G, I) = monitor$ holdsAt $T \wedge$
 $exMethod(I) = discretionary$ holdsAt T
 $pow(G, exclude(G, A, member, I)) = true$
holdsAt $T \leftarrow$
 $role_of(G, I) = monitor$ holdsAt $T \wedge$
 $exMethod(I) = (jury, WDM)$ holdsAt $T \wedge$
 $ballot(exclude(A), I) = V$ holdsAt $T \wedge$
 $winner_determination(WDM, V, true)$
 $per(G, exclude(G, A, member, I)) = true$
holdsAt $T \leftarrow$
 $role_of(G, I) = monitor$ holdsAt $T \wedge$
 $sanctioned(A, T', I) = true$ holdsAt $T \wedge T' < T$

Note that the *monitor* is empowered to exclude any *member*, but it is only *permitted* to exercise that power when that member has been sanctioned (and, when, the exclusion method is *jury*, only when the vote is in favour of exclusion). This means that when the monitor excludes an agent that agent really is excluded and it has no role in the institution. In a richer setting, an excluded agent could appeal against an invalid use of the power, the *monitor* could be removed from the role, and so on. Here we see the entanglement with the system of retributive justice which was beyond the scope of this experiment.

5.1.4 Experimental results

An experimental testbed was written to simulate the iterated LPG game with multiple clusters and a randomly generated population of agents, whereby bundle of information is associated with each agent. This includes its name, up-time, down-time, initial cluster and role assignment (may be none), and its strategy for the LPG game. This strategy is given by a probability of complying with the rules of the game. Therefore, the contribution $r_{i,t}$ that an agent i makes at time t is given by:

$$r_{i,t} = \begin{cases} Ave_{(t-1)} + rnd \cdot (1 - Ave_{(t-1)}) & \text{if } rnd \geq pc_{i,t} \\ rnd \cdot Ave_{(t-1)} & \text{otherwise} \end{cases}$$

where $pc_{i,t}$ is the probability of i 's compliance at t and $Ave_{(t-1)}$ the average cluster contribution from the last time-point. As a result the contribution is in the interval $[Ave, 1]$

if a random number rnd generated in the interval $[0, 1]$ is greater than the probability of compliance, and is in the interval $[0, Ave]$ otherwise.

The agents update their probability of compliance in the next time-point according to a form of social influence. Letting $|I|$ denote the *headcount* for the number of agents in cluster I and $|I|^+$ denote the number of agents in I which complied in the current round, then:

$$pc_i(t+1) = \begin{cases} pc_i(t) + \alpha \cdot (1 - pc_i(t)) & \text{if } (|I|^+ / |I|) \geq 0.5 \\ pc_i(t) - \beta \cdot pc_i(t) & \text{otherwise} \end{cases}$$

where α and β are globally defined coefficients in $[0, 1]$ determining the rate of positive and negative *reinforcement*, respectively. If a majority of agents complies in the current round, then the likelihood of each one complying in the next round is increased, and vice versa.

The results (reported in Pitt et al. 2011) demonstrated the positive effect of clearly defined boundaries and collective-choice arrangements. It turns out that the optimal strategy is to behave like the majority, i.e. if everyone else is complying, then the cost of exclusion from not complying is greater than the benefits gained from not complying (especially when the exclusion method is jury and the ‘moral majority’ is likely to vote for exclusion). On the other hand, if everyone else is cheating, it is a mug’s game not to cheat as well. In our experiments, it was observed that clusters would form which oscillated around specification points 3 and 4 (i.e. the average compliance was >0.5 , and in fact tended to 1.0); the clusters themselves were stable and enduring; and that compliance pervasion emerged even from low level of initial compliance (even with an initial probability of compliance for all the population of 0.2, clusters eventually formed in which the compliance approached 1.0).

However, in this setting, an agent that did not comply with the provision rule was met with one possible sanction, exclusion. This presupposes that provision is observable, and does not prohibit an agent from re-joining a cluster from which it has been excluded. A finer-grained analysis of the provision and appropriation actions is required to allow for both intentional and unintentional errors, and a correspondingly finer-grained system of sanctions. For this, policies for retributive justice are required.

5.2 Retributive justice

5.2.1 Experimental setting

In our view, the LPG game makes several ‘unrealistic’ assumptions, including:

- there in the general setting of an open system, there is no *full disclosure* of agents internals to check compliance;
- there is no capacity to cheat on appropriation so that what is allocated is what is appropriated: however, agents can ‘cheat’ on appropriation by taking more than they are allocated;
- it is assumed that monitoring costs are essentially for ‘free’, but in a computational system with endogenous resources the cost of monitoring (and indeed of computing the resource allocation) has to be ‘paid for’ using the same pool of resources as those to be allocated.

Therefore, to study retributive justice in self-organising open systems, we have proposed a variant game, *LPG'*.

In an *LPG'* game, n agents form a *cluster*. The game itself is played in consecutive rounds, $t_0, t_1, \dots, t_\infty$ (we omit identifying a round by a subscript t if it is clear from context). In each round t , each (player) agent i :

- Determines the resources it has available, $g_i \in [0, 1]$.
- Determines its need for resources, $q_i \in [0, 1]$ ($q_i > g_i$).
- 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, $r'_i \in [0, 1]$.

The *LPG'* game assumes an *economy of scarcity*, by insisting that each agent’s need for resources is greater than those it is capable of generating for itself, i.e. $q_i > g_i$ for each agent i . Thus, agents are necessarily dependent on others, there is an incentive not to comply with the rules, and the resource allocation method has to cope with the deficiency.

The total resources accrued by an agent at the end of a round is given by R_i :

$$R_i = r'_i + (g_i - p_i)$$

i.e. R_i is the sum of the resources that are *appropriated* (rather than allocated) from the common-pool and the available resources that are withheld from the pool. The utility of agent i is then given by:

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

where a , b and c are coefficients in \mathbb{R} measuring, respectively, the relative utilities of getting resources that are needed, getting resources that are not needed, and not getting resources that are needed. Note that appropriated resources do not accrue from one round to the next.

Independent of its utility, each agent i in cluster C makes a *subjective* assessment of its *satisfaction* $\sigma_{i,C}$, represented as a value in $[0,1]$, based on its allocation in relation to its demands. Each agent increases its satisfaction in the next round if it is allocated at least the same as its demand in the current round, and decreases it otherwise:

$$\sigma_{i,C}(t+1) = \begin{cases} \sigma_{i,C}(t) + \alpha \cdot (1 - \sigma_{i,C}(t)) & \text{if } r_i \geq d_i \\ \sigma_{i,C}(t) - \beta \cdot \sigma_{i,C}(t) & \text{otherwise} \end{cases}$$

where α and β are coefficients in $[0,1]$ which determine the rate of reinforcement of, respectively, satisfaction and dissatisfaction. We also define a threshold or cut-off value τ and an interval value m such that if for m consecutive rounds, an agent i evaluates $\sigma_i(C, t) < \tau$ as true, then it will leave the cluster C . It will not re-join a cluster it has previously left.

We note, in passing, that this representation of ‘satisfaction’ is a rudimentary form in interactional justice, i.e. a subjective assessment of an individual’s treatment in the context of an institution.

5.2.2 Formal characterisation

The resource allocation rule for C is *Ration*⁺. This is a mapping from an (indexed) set of demands (for resources) by the agents in C to an (indexed) set of allocations, given by:

$$Ration^+ : \{d_i\}_{i \in C} \rightarrow \{r_i\}_{i \in C}$$

Three further operational-choice rules are designed to regulate the agents’ participation in C . These are:

1. *Provision* rule—an agent must provide what it has available:
 $ocr_p : \forall i \in C, p_i = g_i$
2. *Appropriation* rule—an agent must not appropriate more than it is allocated:
 $ocr_a : \forall i \in C, r'_i \leq r_i$
3. *Moderation* rule—an agent must not demand more than it needs:
 $ocr_m : \forall i \in C, d_i \leq q_i$

Of the six values involved in these three rules, we can distinguish between:

- Two *internal* values, g_i and q_i , whose values, we assume, cannot be determined except by disclosure or an audit of agent i ’s local state.

- One *externalised* value, d_i , whose value is determined individually and disclosed (as an institutional fact) by each empowered agent i (in the role of *prosumer* in cluster C).
- One *computed* value, r_i , which is an institutional fact whose value is determined by the resource allocation rule and initiated by one empowered agent, h , who occupies the role of the head of the cluster.
- Two *physical* values, p_i and r'_i , which can be asserted as institutional facts, either objectively by monitoring the corresponding provision and appropriation actions (at a certain cost); or subjectively by disclosing the corresponding action (which we assume is verified ‘for free’).

Assuming that the physical values can be monitored but the internal values cannot be audited, then it is possible to monitor an agent’s provision but impossible to verify compliance with the rule. Therefore, we only have one monitoring function for the appropriation action:

$$monitor_ocr_a : \forall i \in C, (i, r_i, r'_i) \rightarrow Boolean$$

The monitor rule takes an agent, its allocation and its appropriation, and returns *true* if the appropriation is larger than or equal to the allocation (i.e. the agent has been caught ‘cheating’), and *false* otherwise. Related to the monitoring rule is the cost of applying the rule, which will be used to inform a decision about the frequency of monitoring, up to a certain maximum level. (Recall monitoring has to be ‘paid for’ from the same resources that are being appropriated.) These values, like those for the graduated sanctions, are determined by collective-choice rules.

5.2.3 Policy specification

The policy specification for this action situation is, as before, given by a set of domain-dependent EC axioms which define the institutionalised powers of the agents to initiate or terminate institutional facts. (In this specification, we use a different representation of *role*. There are multiple ways of representing role and the choice of representation depends on whether agents belong to multiple clusters, whether they can have multiple roles, etc. There is nothing contingent on this.)

In this action situation, there is scope for intentional or unintentional error; therefore, the *report* action is performed if an agent detects a violation. This occurs when it is unable to collect its full appropriation, and is inevitable if some other agent violates the appropriation rule by taking more than it was allocated. An agent is empowered to report a violation, if it is a *prosumer* in the cluster, and

secondly if it appropriated less than it was allocated (because the common pool was exhausted):

$$\begin{aligned}
 &report(I, Rule, G, C) \text{ initiates} \\
 &\quad reported(I, Rule, G, C) = true \text{ at } T \leftarrow \\
 &\quad \mathbf{pow}(I, report(A, Rule, G, C)) = true \text{ holdsAt } T \\
 &\mathbf{pow}(I, report(I, Rule, G, C)) = true \text{ holdsAt } T \leftarrow \\
 &\quad current_game(C) = G \text{ holdsAt } T \wedge \\
 &\quad allocated(I, G, C) = R_i \text{ holdsAt } T \wedge \\
 &\quad appropriated(I, G, C) = R'_i \text{ holdsAt } T \wedge \\
 &\quad R'_i < R_i \wedge \\
 &\quad reported(I, Rule, G, C) = false \text{ holdsAt } T \wedge \\
 &\quad role(I, prosum, C) = true \text{ holdsAt } T
 \end{aligned}$$

The axiomatic specification can be directly expressed in Prolog to give an executable specification. At the end of one round of the *LPG'* game, it is used to compute the institutionalised powers of the agent occupying the role of *head*. For example, consider the following narrative (extract) comprising actions performed by agents g , h and p , occurring between $T = 1,610$ and $T = 1,678$ in the 21st round of the *LPG'* game played in the cluster called *cluster1* (note that the syntactic sugar of the specification the infix *happensAt* predicate has been replaced by the prefix *happens*):

```

happens(demand(p,0.4040986,21,cluster1),1610).
happens(demand(g,0.93242,21,cluster1),1622).
happens(allocate(h,p,0.4040986,21,cluster1),1631).
happens(allocate(h,g,0.409146,21,cluster1),1646).
happens(disclose(g,0.2780482,21,cluster1),1670).
happens(monitor(h,p,0.53519,21,cluster1),1675).
happens(report(g,ocr_a,21,cluster1),1677).
happens(gameover(h,21,cluster1),1678).
    
```

Then the Prolog *findall* query returns the following results:

```

?- findall(PAction, holdsAt(pow(h, PAction)=true, 1679), IP).
IP=[increment(h, ocr_a, 21, cluster1), sanction(h, p, ocr_a, cluster1)].
    
```

This means that in *cluster1*, agent h occupies the role of *head* and is empowered both to increase the monitoring frequency of the *ocr_a* rule (as reported by agent g), and to sanction agent p which violated the *ocr_a* rule.

The fluents with which we were particularly concerned for the system of retributive justice were:

Fluent	Dom	Description
$maxmf(C)$	\mathbb{N}^0	Maximum monitoring frequency of C
$mf(C)$	\mathbb{N}^0	Current monitoring frequency of C
$mfct(G, C)$	\mathbb{N}^0	Number of <i>monitor</i> actions performed during G
$decmf_ctr(C)$	\mathbb{N}^0	Number of rounds since last reported violation

For any cluster C , $maxmf(C)$ was fixed, but $mf(C)$ could be incremented or decremented by an empowered agent (e.g. h in the narrative above) performing the appropriate action to increment (or decrement) this value, for example:

$increment(I, Rule, G, C)$ initiates

$$mf(C) = F1 \text{ at } T \leftarrow$$

$\mathbf{pow}(I, increment(A, Rule, G, C)) = true \text{ holdsAt } T$

$mf(GC) = F \text{ holdsAt } T \wedge$

$maxmf(C) = M \text{ holdsAt } T \wedge$

$F < M \wedge$

$F1 = F + 1 \wedge$

$role(I, prosum, C) = true \text{ holdsAt } T$

This meant that when a violation was reported, the head agent could increase the number of monitoring actions performed in any game round G , up to the maximum monitoring frequency $maxmf(C)$. If no violations were reported for a certain number of rounds, the same agent was empowered to decrement the number of monitoring actions.

5.2.4 Experimental results

A variation of the previous testbed was written to investigate self-organisation of the system of retributive justice in the context of the LPG' game. There is one cluster playing iterated rounds of the game. Each round follows the steps outlined previously. In each round, each agent generates (privately) its available and needed resources and declares (publically) its provided and demanded resources. Each agent now had a new parameter $pCheat$: with probability $pCheat$ an agent will provide fewer resources than it has available (i.e. it violates rule ocr_p , although the violation cannot be detected). The resources are pooled, the monitoring costs are subtracted, and the allocation computed using a rationing rule (each agent is allocated an equal share; if its share exceeds its demand the excess is re-allocated equally to those that have received less, and so on).

The findings of these experiments are reported in (Pitt and Schaumeier 2012). The maximum monitoring

frequency was required, otherwise if cheating was sufficiently frequent then the agents would increase the current monitoring frequency uncontrollably, and exhaust all their pooled resources on monitoring, leaving nothing left to allocate. However, the optimal values of maximum monitoring frequency and sanctioning method critically depended on the profile of the agent population: excessive monitoring resulted in unnecessary costs and overall it would have been better to tolerate a ‘low level’ of non-compliance; however, stronger sanctions are not necessarily better for a generally compliant population—it seems that it is as easy to wreck a self-organising system by excessive monitoring and over-strict sanctioning as it is by non-compliant behaviour. This replicates results on enforcement in open systems reported in (Balke et al. 2013). Both these studies show that specific decisions need to be made about the retributive justice system, but that this system should be subject to the same collective-choice processes (i.e. participation in election by those affected by them) and that it also needs to be congruent with the environment—and the environment includes the behaviour of the agents themselves.

However, these experiments also indicated another direction of inquiry. The resource allocation was computed by the rule $Ration^+$, but this also ignores the behaviour of the agents. Alternative ways of allocating resources should be considered, and for this a policy for distributive justice is required.

5.3 Distributive justice

5.3.1 Formal characterisation

The experimental setting for our experiments in distributive justice is the LPG' game, as introduced above, but with the assumptions of an economy of scarcity and the possibility of cheating on appropriation. We also assume that agents are expected to maximise collective benefit, in that none of them can satisfy their demands on their own, so they have to pool resources for at least some of them to have a chance of satisfying their demands. Then the ‘paradoxical’ nature of the LPG' game is then similar to the LPG . That is, the incentive is to withhold provision, misrepresent need, and appropriate more than is allocated, thereby maximising individual utility and satisfaction. But if, all agents behave like this, then each agent is reduced to relying on its own available resources, satisfaction decreases, and eventually the cluster disintegrates as agents leave the cluster through dissatisfaction.

Therefore, it is necessary to design a mechanism to incentivise provision, encourage accurate representation of needs, and discourage excess appropriation. We do so by defining an allocation mechanism using Rescher’s model

Table 2 Rescher’s canons of distributive justice

1	Treatment as equals
2	Treatment according to their needs
3	Treatment according to their actual productive contribution
4	Treatment according to their efforts and sacrifices
5	Treatment according to a valuation of their socially useful services
6	Treatment according to supply and demand
7	Treatment according to their ability, merit or achievements

of distributive justice, based on the canon of legitimate claims (Rescher 1966) within a self-organising electronic institution. This takes into account the quantity of resources to be allocated, the method by which resources are to be allocated, the outcome of the method as a mapping from the members of the institution to an actual allocation, and the monitoring of their behaviour.

Rescher’s analysis of distributive justice concludes that the Principle of Utility, taken as a fairness metric expressed as “the greater good of the greater number”, is but one of many prevailing considerations which need to be taken into account when determining a ‘fair’ allocation of resources. Rescher then argued that an adequate theory of distributive justice requires coordination of the concepts of justice, construed in terms of fairness and equity, and of utility, in the sense of general welfare.

Rescher’s canons are summarised in Table 2, as the canons of equality, need, ability, effort, supply and demand, productivity and social utility. As discussed earlier, Rescher’s analysis required determining, in context, what the legitimate claims are, how they are accommodated in case of plurality, and how they are reconciled in case of conflict.

In the context of the *LPG’* game, we can specify the legitimate claims as functions that compute a total order over the set of members, each of which determine the relative merit of the member’s claims. For example, *The canon of equality*, can be represented by ranking the agents in increasing order of their average allocations:

$$f_1 : \frac{\sum_{t=0}^T r_i(t)}{T_{\{i \in C\}}}$$

where T denotes the total number of rounds of the *LPG’* game played in a particular cluster C , and $T_{\{i \in C\}}$ denotes the number of rounds that agent i has played in cluster C .

Defining an appropriate function for each canon of legitimate claims deals with the question of representing legitimate claims. To accommodate multiple claims, each canon f_* is treated as a *voter* in a Borda count protocol. Under Borda count voting, each vote ranks the list of

candidates in order of preference. Borda points are assigned to each candidate in the list: for example, with n candidates, rank k scores $n - k + 1$ Borda points. Borda points from each vote are summed to give a total Borda score for each candidate.

In our case, each voting function f_* rank orders all the agents in a candidate list according to the relative grounds for their claims, and the Borda score for each agent is computed from the accumulation of Borda points associated with each vote. Normally, in a Borda count protocol, the candidate with the highest Borda score wins, but we may have multiple ‘winners’, so we form a Borda point queue in descending order of Borda score, and allocate resources to the front of the queue until there are no more to allocate.

To reconcile conflicts that may arise between multiple claims, a *weight* $w_* \in [0, 1]$ is attached to each function $f_* \in F$. The Borda score B of agent i under a set of functions F is given by:

$$B(i, F) = \sum_{*=1}^{|F|} w_* \cdot bpts(f_*(i))$$

where $f_*(i)$ computes the rank order assigned to agent i by each f_* , $bpts()$ computes the Borda points for that rank, and w_* is the weight attached to the corresponding function. At the end of each round, the agents then self-organise the weights on the claims.

5.3.2 Policy specification

The *head* agent then applies the voting functions for legitimate claims, and declares the result as the value of the fluent *Borda_ptq*:

$$\begin{aligned} &declare(H, L, G, C) \text{ initiates} \\ &\quad \text{borda_ptq}(G, C) = L \text{ at } T \leftarrow \\ &\text{pow}(H, declare(H, L, G, C)) = true \text{ holdsAt } T \\ &\text{pow}(H, declare(H, L, G, C)) = true \text{ holdsAt } T \leftarrow \\ &\quad \text{current_game}(C) = G \text{ holdsAt } T \wedge \\ &\quad \text{borda_ptq}(G, C) = [] \text{ holdsAt } T \wedge \\ &\quad \text{role}(H, C) = head \text{ holdsAt } T \end{aligned}$$

An allocation of resources to an agent is initiated by the empowered agent (the agent appointed to the role of *head* of the cluster) performing the designated action. This initiates new values for three fluents: the allocation r_i to the agent, the reduced (institutional) pool *ifpool* of resources to allocate, and removal of agent from the front of *borda_ptq*.

$allocate(H, I, R_i, G, C)$ initiates
 $allocated(I, G, C) = R_i$ at $T \leftarrow$
 $\mathbf{pow}(H, allocate(H, I, R_i, G, C)) = true$ holdsAt T

$allocate(H, I, R_i, G, C)$ initiates
 $ifpool(G, C) = P'$ at $T \leftarrow$
 $\mathbf{pow}(H, allocate(H, I, R_i, G, C)) = true$ holdsAt $T \wedge$
 $ifpool(C) = P$ holdsAt $T \wedge$
 $P' = P - R_i$

$allocate(H, I, R_i, G, C)$ initiates
 $borda_ptq(G, C) = Tl$ at $T \leftarrow$
 $\mathbf{pow}(H, allocate(H, I, R_i, G, C)) = true$ holdsAt $T \wedge$
 $borda_ptq(G, C) = [I \mid Tl]$ holdsAt T

The *head* has the power to allocate to any agent, provided a vote has been taken, but is only *permitted* to allocate to the agent in the front of the queue.

$\mathbf{pow}(H, allocate(H, I, R_i, G, C)) = true$ holdsAt $T \leftarrow$
 $current_game(C) = G$ holdsAt $T \wedge$
 $role(H, head, C) = true$ holdsAt T
 $\mathbf{not} borda_ptq(C) = []$ holdsAt T

$\mathbf{per}(H, allocate(H, I, R_i, G, C)) = true$ holdsAt $T \leftarrow$
 $\mathbf{pow}(H, allocate(H, I, R_i, G, C)) = true$ holdsAt $T \wedge$
 $borda_ptq(G, C) = [I \mid Tl]$ holdsAt T

5.3.3 Experimental Results

A testbed for experimenting with distributive justice in the *LPG'* game was implemented. With a randomly generated population of compliant and non-compliant agents, we experimented with a single cluster with different subsets of the legitimate claims, and multiple clusters with different allocation schemes.

We used a set *F* of eight functions to represent the legitimate claims, three for different aspects of treatment as equals, and one each for Canons 2–6. Canon 7, treatment according to merit, achievement, rank etc. is not relevant in this context.

To evaluate the fairness of the decision arena for cluster *C* for a given run, we used the Gini inequality index over:

$$\left\{ \sum_{i=0}^T allocate_{ij} / \sum_{i=0}^T demand_{ij} : j \in 1 \dots |C| \right\}$$

An index of 0 is perfect equality (i.e. fairness), and 100 is complete inequality.

The general findings were that:

- Only set *F* of all canons with self-organisation produced an enduring cluster which benefited the compliant agents;
- Set *F* of all canons with self-organisation was fairest (wrt. strict ration and all canons with no self-organisation) using the selected fairness measure;
- Compliant agents preferred to join a cluster with resource allocation determined by set *F* of all canons with self-organisation;
- With set *F* of all canons with self-organisation, using the selected fairness metric, a series of ‘rather unfair’ allocations led to a ‘very fair’ overall allocation.

5.4 Summary

With these experiments, we were intending to demonstrate that concepts of justice are critical to formalising Ostrom’s design principles for self-organising electronic institutions, to explore the ‘design space’ underlying each principle and to expose the inter-connectedness and inter-dependence of the principles. However, we have been primarily concerned with rules of operational choice and rules of collective choice. We believe that to address issues of scalability and the mechanisms of representation that for a corresponding ‘economy of scale’, the rules of constitutional choice are required, but this is a matter for further experimental investigation. In the next section, we consider additional directions for further research.

6 The pursuit of computational justice

In this section, we consider the relationship of different aspect of justice to Ostrom’s institutional design principles, highlight several areas for further work and some other open questions, and conclude with a proposal a broader programme of research, *the pursuit of computational justice*.

6.1 Ostrom’s principles and qualifiers of justice

The experimental results offer some insight into the relation of Ostrom’s principles to specific qualifiers of justice, and the inter-dependence of the qualifiers of justice. For example, Ostrom’s Principles (1) clearly defined boundaries and (3) collective-choice arrangements indicate a need for a system of natural justice. Similarly, Ostrom’s Principles (2) refers to appropriation and provision rules, and the need for a system of distributive justice which is both ‘fair’ and ‘stable’, and its selection and modification through Principle (3), is essential. Additionally, Ostrom’s

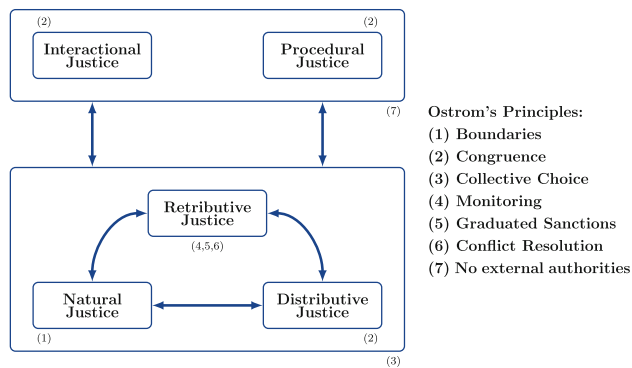


Fig. 2 Inter-dependence of the qualifiers of justice and Ostrom's principles

Principles (4) monitoring, (5) graduated sanctions and (6) access to conflict-resolution procedures indicate a need for a system of retributive justice. However, the institution members should also participate in the selection of the rules embodying these principles as well. As a result, it can be seen that these qualifiers of justice are entangled and inter-dependent.

Furthermore, fully ensuring the *congruence* of the appropriation and provision rules to the state of the prevailing environment indicates a requirement for a system of procedural justice underpinning Principle (2). Additionally, to be able to express a preference in any collective-choice arrangement, the notion of 'satisfaction' needs to be related to individual preferences and social capital, especially in self-organising socio-technical systems. This goes beyond quantitative measures of utility, and requires a computational theory of emotion together with a system of interactional justice, and in fact the first part—the need for interpersonal justice subjectively measured by 'quality of treatment'. Thus, both these qualifiers of justice are in some way concerned with 'evaluating' the inter-dependent system of natural, distributive and retributive justice.

This relationship between the qualifiers of justice and how they help encapsulate Ostrom's principles is illustrated in Fig. 2.

In other words, if we encode a policy for each qualifier of justice, then we are addressing each of Ostrom's first six principles. It has been shown that—assuming there is no external interference and a single system (Principles 7 and 8)—that these are also essential and determinate conditions for self-organising electronic institutions (Pitt et al. 2012).

6.2 Further work

It is clear that we have barely scratched the surface of formal characterisation and principled operationalisation of systems of procedural justice and interactional justice, but this is work in progress. Some other issues for further research

concern applying Ostrom's principles and computational justice to knowledge commons, visualisation in socio-technical systems, evidence-based policy-making and self-organising learning agents. We briefly consider each in turn.

6.2.1 Knowledge commons

The work presented in (Ostrom and Hess 2006) was concerned with treating knowledge as a shared resource, motivated by the increase in open access science journals, digital libraries, and mass-participation user-generated content-management platforms. It then addressed the question of whether it was possible to manage and sustain a knowledge commons, using the same principles used to manage ecological systems with natural resources. A significant challenge in the democratisation of Big Data is the extent to which formal representations of intellectual property rights, access rights, copy-rights, etc. of different stakeholders can be represented in a system of computational justice and encoded in Ostrom's principles for knowledge commons. As observed in (Shum et al. 2012), the power of Big Data and associated tools for analytical modelling: "... should not remain the preserve of restricted government, scientific or corporate élites, but be opened up for societal engagement and critique. To democratise such assets as a public good, requires a sustainable ecosystem enabling different kinds of stakeholder in society".

6.2.2 Visualisation

People occupying a physical space have to access its physical resources and services water, energy, mobility, etc. To this end, an infrastructure is developed and deployed; the automation of that infrastructure results in calling it 'Smart' (SmartGrids, SmartCities, etc.), and demand-side active participation and user engagement is presumed. However, the user-infrastructure interface and interaction dynamics are often neglected with critical consequences for sustainability (Lam 1998). To get both a better understanding of the behaviour of energy consumers and getting energy consumers to understand better the effects of their behaviour on the grid, virtual environments can be used as an innovative energy infrastructure interface, with a particular emphasis on visualisation of the principles for self-organisation of the resource allocation (Bourazeri et al. 2012). Effectively, we are looking towards the visualisation, animation and explanation of the justice system(s) underlying all decision-making in the 'Smart' infrastructure.

6.2.3 Evidence-based policy-making

This is another issue concerning socio-technical self-organising systems. Whatever policies are developed in

order to achieve macro-level goals, like sustainability, or low-carbon emission, etc., people do not simply comply or not comply with the policy, they react to incentives implied by the policy (López 2010). Those incentives often stimulate behaviour different from what was intended, because people have other incentives to find loopholes in the policy or because the policy has unintended secondary and even tertiary consequences. The logical representation of policy as a system of computational justice, in conjunction with agent-based population simulation, offers an intriguing route to building an evidence-base for informing systematic decision-making.

6.2.4 Self-organising learning agents

A major challenge to build durable open systems holds in the cybernetic loops between norm-governed agents and institutions. To adapt to and learn to shape their normative environment, a solution consists in endowing autonomous agents with reinforcement learning. Considering rewards and punishments, they may comply and internalise the cost of non-compliance. As the social system may turn to be unfair, these agents shall also learn to influence existing institutions and even self-organise to constitute their own. In this view, and in conjunction with an extractable algebraic representation akin to utilitarian quantitative calculus of implementation theory, the executable logical representation of norm-governed learning agents as proposed in (Riveret et al. 2012) paves the way for the study of durable self-organising and self-managing systems of learning agents.

6.3 Clearly defined boundaries?...

In this paper, we have considered the problem of resource provision and appropriation in various types of open ‘technical’ and ‘socio-technical’ system, identified a set of key features common to these systems, proposed ‘justice’ as the social concept whose various qualifiers provide a basis for addressing these features, and then formalised (qualified notions of) justice within the institutional design principles of Elinor Ostrom. However, this programme of work has entailed various ‘design decisions’, and it is appropriate to reflect on the rationale for those decisions, as it too opens up a number of research questions.

There are three reflections which we consider here. The first reflection is: are the six features identified in Sect. 2.3 the essential and determinate conditions for resource allocation in open socio-technical systems? To answer this question conclusively, a more thorough and systematic characterisation and classification of open systems is required, to identify those which do exhibit the features which can be addressed by the systems of justice proposed

here, and those which perhaps cannot. It might be that further analysis could reveal that there are yet more features to address, which are specifically relevant to resource allocation; it might also reveal whether or not these features—and others—are relevant to addressing additional open system functionalities other than resource allocation.

The second reflection has three parts: is justice the essential and determinate concept for addressing these features? Are the five qualifiers identified here the essential and determinate set for solving resource allocation problems? And, is justice, and the five qualifiers identified here, the essential and determinate concept for solving collective action problems of this kind in general? We have proposed justice as a key concept in resource provision and appropriation, but *trust* has also been identified as the ‘glue’ which links social capital to solving collective action problems (Ostrom and Ahn 2003). The relationship between formal frameworks of trust (and its counterpart, forgiveness) in the administration of justice is an interesting open question.

The third reflection is: are Ostrom’s theories the essential and determinate conditions for principled operationalisation of the kind practised in this paper? We have identified Ostrom’s work as the key source from the social sciences used in our formalisation, because, (methodologically) faced with an engineering challenge (resource allocation in open systems) we asked “how do people solve this problem?“, and *one* answer is given by Ostrom. There may well be other answers; it may well be that Ostrom’s sources do not provide all the answers (indeed we ‘instantiated’ Ostrom’s principles with the theory of distributive justice from Rescher). We need to be clear that nothing is ruled out; but if the answer is rooted in the social sciences we would advocate using the methodology for design of intelligent systems proposed in (Jones et al. 2013). It is in this sense that ‘principled-ness’ is derived: understanding what may be lost from a theory in process of formal characterisation, and understanding what might be ‘added’ to the underlying theory because of specific engineering constraints and requirements.

6.4 Beyond resource allocation...

The initial study of resource allocation, in the context of open self-organising systems (Pitt et al. 2012), has developed into a rich and diverse research programme which can be characterised as *computational justice*, an inter-disciplinary investigation at the interface of computer science and philosophy, economics, psychology and jurisprudence. Therefore, computational justice is an interdisciplinary study at an intersection between computer science and social sciences, enabling and promoting an exchange of ideas and results in both directions.

From one perspective, computational justice is concerned with the application of formal representations of and reasoning about justice developed in computer science. For example, the analysis of the free-rider paradox in peer-to-peer file-sharing systems has significant implications in certain social settings, in particular, for e-commerce models based on incentives and voluntary contributions. It also concerned with making the outcomes of this representation and reasoning conceptually and perceptually clear to users of socio-technical systems for managing social infrastructure, through visualisation, explanation, active participation, and so on. From the other perspective, it is concerned with importing concepts from the social sciences into computing applications, for example, by implementing theories of procedural, distributive or retributive justice in multi-agent systems.

From both perspectives, computational justice has much to offer the study of self-organising systems. The objective of this paper is to contextualise computational justice in self-organising electronic institutions.

Furthermore, computational justice is not restricted to resource allocation, but many other areas of technical and socio-technical concern, including online-dispute resolution (Katsh et al. 2001) and indeed the fairness concerns of computational social choice (Chevalleyre et al. 2007).

7 Summary and conclusions

The argument we have made in this paper is to have started from a review of open systems, both ‘technical’ and socio-technical, and identified a set of common problematic features. We then reviewed different qualifiers of justice and aligned each qualifier of justice to a problematic feature. We then presented a formal model, self-organising electronic institutions, in which a system of *-justice could be represented in computational form. Using this formal model, we presented the results of three experiments with systems of natural, retributive and distributive justice.

Our basic ambition was to implement each of Elinor Ostrom’s principles for enduring self-governing institutions in computational form. However, each successive experiment revealed that there were multiple inter-dependent influences at work, and each of these influences could be attributed to the requirement for some form of justice. The experimental results indicate that self-organisation in open systems requires the representation and principled operationalisation of all these inter-dependent systems of justice. These experiments alone did not answer all the questions we raised about fairness, efficiency, accountability, etc. that we raised at the start and also exposed some new questions and other lines of research. Furthermore, while our work has been heavily influenced by

Ostrom’s theories, there are many other intriguing theories of justice that could provide a basis for formalisation (e.g. (Rawls 1971)), and the notion of agency, structure and function could be explored at a more fundamental level, for example, using the theory of structuration of Giddens (1984).

As a result, we have tried to open out this work to propose a broader programme of research we have called *computational justice*. As Sect. 6 indicates, there is still much to do.

Acknowledgments We are particularly grateful for the constructive comments of the four anonymous reviewers. The second and third authors both hold Marie Curie Intra-European Fellowships.

References

- Ardagna D, Panicucci B, Passacantando M (2011) A game theoretic formulation of the service provisioning problem in cloud systems. In: *www '11*, pp 177–186
- Artikis A (2012) Dynamic specification of open agent systems. *J Logic Comput* 22(6):1301–1334
- Balke T, de Vos M, Padget J (2013) I-ABM: combining institutional frameworks and agent-based modelling for the design of enforcement policies. *Artif Intell Law*. doi:10.1007/s10506-013-9143-1
- Bentham J (1789) *An Introduction to the principles of morals and legislation*. Payne, London
- Bernard Y, Klejnowski L, Cakar E, Hähner J, Müller-Schloer C (2011) Efficiency and robustness using trusted communities in a trusted desktop grid. In: *SASO Workshops (SASOW)*, pp 21–26. doi:10.1109/SASOW.2011.28
- Binmore K (2005) *Natural justice*. Oxford University, Oxford
- Birman K, Chockler G, van Renesse R (2009) Toward a cloud computing research agenda. *SIGACT News* 40(2):68–80
- Blader A, Tyler T (2003) A four-component model of procedural justice: defining the meaning of a “fair” process. *Pers Soc Psychol Bull* 29(6):747–758
- Bourazeri A, Pitt J, Almajano P, Rodriguez I, López-Sánchez M (2012) Meet the meter: visualising smartgrids using self-organising electronic institutions and serious games. In: *SASO Workshops (SASOW)*, pp 145–150
- Chevalleyre Y, Endriss U, Lang J, Maudet N (2007) A short introduction to computational social choice. In: *Proceedings of the 33rd conference on current trends in theory and practice of computer science (SOFSEM)*, no. 4362 in LNCS. Springer, Berlin, pp 51–69
- Choi S, Buyya R, Kim H, Byun E (2008) A taxonomy of desktop grids and its mapping to state of the art systems. *Tech. rep.*, Grid Computing and Distributed Systems Laboratory, The University of Melbourne
- Cox M, Arnold G, Villamayor Tomás S (2010) A review of design principles for community-based natural resource management. *Ecol Soc* 15(4):38
- Ding M, Cheng X, Xue G (2003) Aggregation tree construction in sensor networks. In: *IEEE vehicular technology conference*, vol 4, pp 2168–2172
- Elster J (1992) *Local justice: how institutions allocate scarce goods and necessary burdens*. Russell Sage Foundation, New York
- Ferscha A, Davies N, Schmidt A, Streitz N (2011) Pervasive socio-technical fabric. *Procedia Comput Sci* 7:88–91

- Gaechter S (2006) Conditional cooperation: behavioral regularities from the lab and the field and their policy implications. Discussion Papers 2006-03, The Centre for Decision Research and Experimental Economics, School of Economics, University of Nottingham
- Giddens A (1984) *The constitution of society: outline of the theory of structuration*. Polity, Cambridge
- Greenberg J (1993) The social side of fairness: interpersonal and informational classes of organizational justice. In: Cropanzano R (ed) *Justice in the work place: approaching fairness in human resource management*, Erlbaum, Hillsdale, NJ, pp 79–103
- Hardin G (1968) The tragedy of the commons. *Science* 162(3859):1243–1248
- Hewitt C (1986) Offices are open systems. *ACM Trans Office Inf Syst* 4:271–287
- Jones A, Artikis A, Pitt J (2013) The design of intelligent socio-technical systems. *Artif Intell Rev* 39(1):5–20
- Jones A, Sergot M (1996) A formal characterisation of institutionalised power. *J IGPL* 4(3):427–443
- Kass N (2001) An ethics framework for public health. *Am J Public Health* 91(11):1776–1782
- Katsh EE, Katsh ME, Rifkin J (2001) *Online dispute resolution: resolving conflicts in cyberspace*. Wiley, New York, NY
- Konow J (2003) Which is the fairest one of all? A positive analysis of justice theories. *J Econ Lit* 41(4):1188–1239
- Kowalski R, Sergot M (1986) A logic-based calculus of events. *New Gener Comput* 4:67–95
- Lam WF (1998) *Governing irrigation systems in nepal: institutions, infrastructure and collective action*. ICS, Oakland, CA
- López E (2010) *The pursuit of justice: law and economics of legal institutions*. Palgrave MacMillan, New York, NY
- Manvi SS, Kakkasageri MS, Pitt J (2009) Multiagent based information dissemination in vehicular ad hoc networks. *Mobile Inf Syst* 5(4):363–389
- Nozick R (1974) *Anarchy, state, and utopia*. Basic Books, New York, NY
- Ostrom E (1990) *Governing the commons*. CUP, Cambridge
- Ostrom E, Ahn TK (2003) Introduction. In: Ostrom E, Ahn TK (eds) *Foundations of social capital*, Edward Elgar, Cheltenham, pp xi–xxxix
- Ostrom E, Hess C (2006) A framework for analyzing the knowledge commons. In: Hess C, Ostrom E (eds) *Understanding knowledge as a commons: from theory to practice*, MIT, Cambridge, MA, pp 41–82
- Pitt J, Schaumeier J (2012) Provision and appropriation of common-pool resources without full disclosure. In: *Principles and practice of multi-agent systems*, LNCS, vol 7455, Springer, pp 199–213
- Pitt J, Schaumeier J, Artikis A (2011) Coordination, conventions and the self-organisation of sustainable institutions. In: *Agents in principle, agents in practice*, LNCS, vol 7047, Springer, Berlin, pp 202–217
- Pitt J, Schaumeier J, Artikis A (2012) Axiomatisation of socio-economic principles for self-organising institutions: concepts, experiments and challenges. *ACM Trans Auton Adapt Syst* 7(4):1–39. doi:10.1145/2382570.2382575
- Pitt J, Schaumeier J, Busquets D, Macbeth S (2012) Self-organising common-pool resource allocation and canons of distributive justice. In: *Sixth IEEE international conference on self-adaptive and self-organizing systems (SASO)*, pp 119–128
- Rawls J (1971) *A theory of justice*. Harvard University, Harvard, MA
- Rescher N (1966) *Distributive justice*. Bobbs-Merrill, New York, NY
- Riveret R, Rotolo A, Sartor G (2012) Probabilistic rule-based argumentation for norm-governed learning agents. *Artif Intell Law* 20(4):383–420
- Robert SC, Robert H, Evans WJ, Honemann DH, JBT (2000) *Robert's rules of order, newly revised*, 10th edn. Perseus, Cambridge, MA
- Shum SB, Aberer K, Schmidt A, Bishop S, Lukowicz P, Anderson S, Charalabidis Y, Domingue J, de Freitas S, Dunwell I, Edmonds B, Grey F, Haklay M, Jelasity M, Karpistenko A, Kohlhammer J, Lewis J, Pitt J, Sumner R, Helbing D (2012) Towards a global participatory platform: democratising open data, complexity science and collective intelligence. *Eur Phys J Spec Top* 214:109–152
- Solum L (2004) Procedural justice. *South Calif Law Rev* 78(181):275–289
- Uphsur R (2002) Principles for the justification of public health intervention. *Can J Public Health* 93:101–103