Interdependent Self-Organizing Mechanisms for Cooperative Survival

Matthew Scott*
Imperial College London
Department of Electrical and Electronic Engineering
matthew.scott18@imperial.ac.uk

Jeremy Pitt
Imperial College London
Department of Electrical and Electronic Engineering

Abstract Cooperative survival "games" are situations in which, during a sequence of catastrophic events, no one survives unless everyone survives. Such situations can be further exacerbated by uncertainty over the timing and scale of the recurring catastrophes, while the resource management required for survival may depend on several interdependent subgames of resource extraction, distribution, and investment with conflicting priorities and preferences between survivors. In social systems, self-organization has been a critical feature of sustainability and survival; therefore, in this article we use the lens of artificial societies to investigate the effectiveness of socially constructed self-organization for cooperative survival games. We imagine a cooperative survival scenario with four parameters: scale, that is, n in an n-player game; uncertainty, with regard to the occurrence and magnitude of each catastrophe; complexity, concerning the number of subgames to be simultaneously "solved"; and opportunity, with respect to the number of self-organizing mechanisms available to the players. We design and implement a multiagent system for a situation composed of three entangled subgames—a stag hunt game, a common-pool resource management problem, and a collective risk dilemma—and specify algorithms for three self-organizing mechanisms for governance, trading, and forecasting. A series of experiments shows, as perhaps expected, a threshold for a critical mass of survivors and also that increasing dimensions of uncertainty and complexity require increasing opportunity for self-organization. Perhaps less expected are the ways in which self-organizing mechanisms may interact in pernicious but also self-reinforcing ways, highlighting the need for some *reflection* as a process in collective self-governance for cooperative survival.

Keywords
Artificial societies, multiagent systems, self-organization, cooperative survival games

1 Introduction

Cooperative survival "games" are situations in which, following a catastrophic event, no one survives unless everyone survives. Elements of such situations feature prominently in computer games and board games but have been more seriously and extensively studied in sociology and anthropology, especially with regard to societies surviving in difficult or hostile environments (e.g., Briggs, 1970; Norberg-Hodge, 1991). Similar features appear in embedded cyberphysical systems,

* Corresponding author.

such as ad hoc networks and sensor networks (Durmus et al., 2011); evolutionary systems needing to reconfigure themselves after a dramatic change in their operating environments (Pitt & Hart, 2017); and sociotechnical systems, such as community energy systems, relying on renewable energy, where attention and collective action may be required to prevent a generation shortfall leading to a brownout or blackout (Bourazeri & Pitt, 2018).

These situations can be further exacerbated by uncertainty over the timing and scale of the catastrophe (Domingos et al., 2020) and the overall game itself being composed of multiple subgames of resource extraction, resource distribution, and resource investment, each with conflicting priorities and preferences (Bekius et al., 2022). For example, in a situation prone to flooding, investing too many resources into flood defenses wastes utility, but investing too few can be fatal or cost more in the long term. When there are multiple codependent communities, there may be multiple interacting common-pool resource (CPR) management problems operating on different timescales. For example, an irrigation system serving multiple communities may have both an appropriation problem and a maintenance problem: If one community appropriates excessively, then the others may not contribute to maintenance, to the detriment of all (Ostrom, 1990). Similar timing and structural problems have been observed in other agricultural settings (Lansing & Kremer, 1993).

In this article, we specify a scenario based on an archipelago of islands, each of which needs the others to survive recurring disasters. We define an *iterated* cooperative survival game composed of three interdependent subgames: a stag hunt game, a CPR management problem, and a collective risk dilemma (CRD). The stag hunt subgame (Carlsson & van Damme, 1993) is an n-player cooperative resource extraction game with limitations: For example, it is possible to exhaust the resource by overhunting. This subgame precedes a resource distribution subgame that is an n-player CPR management problem (Ostrom, 1990) in which the players have to decide, individually and collectively, how much of these hunted resources should be provisioned to, and appropriated from, a common pool. This problem in turn precedes a resource investment subgame that is an n-player CRD (Domingos et al., 2020) in which the players have to decide how much of their appropriated resources are used to mitigate the next catastrophic event. Failure to do so determines their ability to participate effectively in the subsequent iteration. Underpinning these subgames are the goals of the cooperative survival game, namely, not just for all to survive but also to find a mutually acceptable balance between individual and collective goals, for example, between individual cognitive effort and quality of the overall outcome in information processing (Nowak et al., 2020) or between engagement in personally valued pursuits and participation in socially productive duties for self-governance (Ober, 2017).

To address this cooperative survival scenario and its component subgames, we use the lens of artificial societies (Powers, 2018) to design three interdependent, socially constructed self-organizing mechanisms, one each for governance, trading, and forecasting. This is based on the observation that, in previous work, each mechanism has been targeted at one game, with forecasting derisking the CRD (Domingos et al., 2020), governance informing the CPR management (Ostrom, 1990), and gifting for redistributing inequalities and creating a form of social capital (Malinowski, 1920). By *socially constructed*, we mean that these mechanisms are intrasubjective agreements based on networked interaction and communication (Berger & Luckmann, 1966), producing conceptual resources, such as trust, institutions, and reciprocity, that have been shown to enhance opportunities for successful collective action (Ostrom & Ahn, 2003; Petruzzi et al., 2017).

Consequently, the challenge becomes as follows: *Rather than one mechanism solving each subgame separately, what combination of which form of these self-organizing mechanisms "optimally solves" the cooperative survival game under different dimensions of scale, uncertainty, and constraints?*

Accordingly, this article is structured as follows. Section 2 describes the nature of both the cooperative survival games and the means of self-organization. We approach the former from the perspective of utility calculations and the dilemma faced when attempting to tackle these problems. Following this, we describe the system that we propose in section 3 as a means of investigating the problem of cooperative survival as construed here, as well as the codification of the various system elements in section 4.

Section 5 presents a series of experiments of increasing uncertainty over the dimensions of the catastrophe, scale of the system, and constraints on the availability of the different self-organizing mechanisms. The results of these experiments show that this problem is solvable, as long as there is a sufficient number of players and the availability of self-organizing mechanisms is judiciously selected due to the possibility of both mutually reinforcing and pernicious interactions. Furthermore, with increasing levels of uncertainty in the system, increasing opportunities for such self-organizations are required. After a consideration of related and further work in section 6, we conclude in section 7 by arguing that the dependencies between subgames and self-organizing mechanisms exist not only between each other but among themselves as well, in complex ways, highlighting the need for *reflection* as an essential process in collective self-governance for cooperative survival.

2 Cooperative Survival Games and Self-Organization

This section presents the conceptual background to this work, covering cooperative survival games and the parameters of scale, uncertainty, complexity, and opportunity (mechanisms for self-organization). It starts in section 2.1 with an informal description of cooperative survival games from different viewpoints. On this basis, section 2.2 introduces the specific scenario used in this work and establishes the parameters of scale and uncertainty, while section 2.3 establishes the dimension of complexity with a review of three interdependent games: a stag hunt game, a public goods game, and a CRD. Finally, section 2.4 presents the opportunities available to the agents—three self-organizing mechanisms of trading, forecasting, and governance—which essentially create meta-level political games that may defuse an "inevitable tragedy" or change the constraints inherent in an object-level resource competition game (Ostrom, 1990).

2.1 Cooperative Survival Games

An instance of cooperative survival can be found in the genre of computer games in which players must work together to survive in an open, hostile environment. It is also the basic premise of round-based board games like Escape the Dark Castle and Ravine, where a crucial feature of game play is the need to keep enough players alive in any one round to gather sufficient resources to ensure that even the weakest players survive and can participate in resource gathering in the subsequent round.

Moreover, cooperative survival is a phenomenon witnessed in many aspects of "real life." Human beings are a social (and, indeed, cooperative) species that relies on cooperation to survive and thrive (Bowles & Gintis, 2011), and in anthropology, there are many examples of survival even in extremely hostile environments (e.g., Briggs, 1970; Norberg-Hodge, 1991).

However, the capability to overcome CRDs is demonstrated from a young age, where children as young as six years old can spontaneously find ways to collaborate to maintain a shared, limited resource. This was shown in the context of a common pool of resources paradigm involving a shared water source in which children were capable of collectively preventing resource collapse by creating inclusive rules, equally distributing the rewards and distracting one another from the delay-of-gratification task (Koomen & Herrmann, 2018). Furthermore, children can be seen to display notions of fairness from a young age as well, making sacrifices for fairness when they have less than others, when others have been unfair, and when they have more than others. This goes against rational self-interest for the good of cooperation (McAuliffe et al., 2017).

Similar features of cooperative behavior for individual or collective survival can be observed in cyberphysical and sociotechnical systems. In cyberphysical systems, for example, there can be a trade-off between accuracy and longevity in a sensor network (Nikoloska & Simeone, 2021) and between quality of service (QoS) and end-to-end connectivity in an ad hoc network. In a sensor network, to maximize the objective of accuracy, all sensors should be transmitting all the time, but to maximize the objective of longevity, only enough sensors needed for acceptable accuracy should be taking measurements and transmitting signals; ideally, only those sensors with the most resources would measure and transmit, to avoid "exhausting" nodes with fewer resources, which

would thereby threaten the network as a whole. Similarly, in an ad hoc network, to maximize through put, as many nodes as possible should be operative, but to ensure the objective of end-to-end connectivity, only enough nodes needed for acceptable QoS should be forwarding packets. There is considerable interest in the notion of resilience in such systems, particularly with regard to disaster response (Ti et al., 2022).

Elements of cooperative survival can also be seen in some sociotechnical systems, for example, a community energy system (Pitt et al., 2014). Energy self-sufficiency in a community energy system redirects the onus of resource management from the demand side (global generators must match the request made by all consumers) to the supply side (local consumers distribute the available energy generated from their own sources). This converts the problem to sustainable management of a CPR (i.e., a public good) that is both finite and might be insufficient to meet demand, otherwise there will be a brownout or, worse, a blackout. This, then, would seem to create the conditions for a tragedy of the commons: One person alone has no incentive to voluntarily reduce their demand to avoid the outage, because unless everyone else also reduces their demand, the sole volunteer loses utility in the short term and still suffers the same tragedy as everyone else. Therefore there needs to be some intervention, such as self-governing institutions (Ostrom, 1990), which have been shown to enable (human) participants to avoid the supposedly inevitable tragedy of the commons.

It is worth noting that the cooperative survival game Minecraft allows the specification of various mods and plug-ins: It has been observed that sustainable Minecraft hosts have implemented mods and plug-ins that reproduce the features of Ostrom's self-governing institutions (Frey & Sumner, 2018). In other words, cooperative survival can be negotiated even by (anecdotally) most antisocial groups given sufficient incentive and adequate opportunity for social construction of a solution.

2.2 Scenario Specification

On the basis of this general discussion of cooperative survival games, the cooperative survival scenario proposed for this article is illustrated in Figure 1. We imagine an archipelago of islands that regularly have to cooperate to mitigate disaster in the form of a randomly located earthquake with an epicenter limited to the bounds of the archipelago. The location, magnitude, and timing of the earthquake are unknown; therefore how badly a disaster affects each island depends on its distance from the epicenter and on the magnitude and timing. To mitigate the disaster, the islands have to contribute their personal resources to a common pool, which, inconveniently, has the dual role

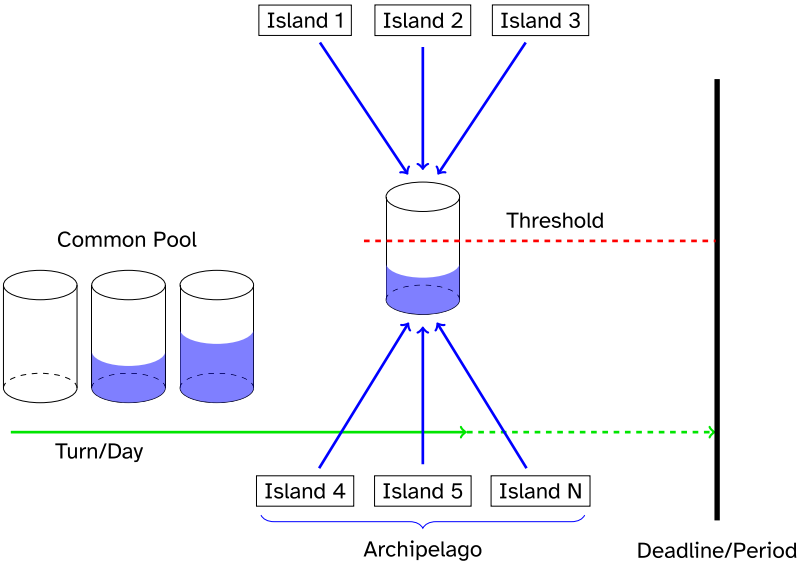


Figure 1. Visualization of collective risk dilemma and archipelago.

of also serving as the communal fund, as islands may withdraw resources from it at will. Furthermore, to provision resources to the CPR, they have to participate in a resource extraction activity, possibly in cooperation with other islands, to forage for either fish, yielding few resources at a low risk, or deer, yielding more resources at a higher risk but requiring coordination between islands. Islands must then decide how much to keep for themselves and how much to provision. If an island's personal resources fall below a critical level for a certain length of time, it will cease to exist (the inhabitants will "die").

The overall scenario can then be decomposed into several subgames: a resource gathering game (i.e., a stag hunt game), a resource distribution game (i.e., a CPR management problem or linear public goods game), and a resource contribution game (i.e., a disaster mitigation game or CRD). Note that the interdependent subgames are designed to create a vicious circle or a virtuous spiral: An inefficient stag hunt yields fewer resources to distribute from the common pool, which provides fewer resources for disaster mitigation and recovery, which can lead to an even more inefficient stag hunt (especially if islands "die"). Conversely, a more efficient stag hunt provides more resources for the common pool, providing more opportunity for disaster mitigation, maintaining critical mass and creating social capital, which leads to a more efficient stag hunt, and so on. A less badly affected island may help out one that is affected worse, helping to ensure its survival in the short term and creating *social capital* for reciprocal behavior in the longer term.

Next, we briefly review each type of game. Note that in the following discussion, an island in the scenario is synonymous with a player in a subgame and is also synonymous with an agent in the multiagent system to be developed in section 3.

2.3 Interdependent Games

2.3.1 Stag Hunt Game

A stag hunt game is an example of a two-player collective choice game that illustrates a conflict between collectivity and individuality. In this game, two players are confronted with two foraging options, for either a stag or a hare (although for contextual reasons, we use deer and fish, respectively, in this article), and must choose which species to forage, both separately and without the other's knowledge. Choosing to forage for a stag will provide maximal rewards for both parties, but only if the other player makes the same decision, else they will get nothing. Alternatively, choosing to forage for a hare will result in lesser returns but the guarantee to receive this amount, as hares can be foraged individually. This article elects the payoff matrix for this game as shown in Figure 2. This formulation preserves the ordinality of the conventional stag hunt game (i.e., $(S,S) > (H,S) >= (H,H) > (S,H)$, for the row player); however, the numerical payoffs result from a combination of the utility functions detailed in section 4.5 and the scaling defined in section 4.6.

For this article, we make several adjustments to the conventional stag hunt dilemma. To begin with, we allow for all N players to play this game with a fixed probability of catching a deer, resulting in $N/2$ stag hunt games being played simultaneously.

		Column Player	
		D	F
Row Player	D	$(48, 48)$	$(0, 26)$
	F	$(26, 0)$	$(26, 26)$

Figure 2. Payoff matrix for resource generation.

In addition, we allow for repeated foraging of either fish or deer by introducing the concept of a *utility tier*, akin to a step function, where more resources can be invested for a stochastic chance of a higher yield, effectively multiplying payoffs (section 4.5.3). Again, this augments the conventional utility calculation to incorporate the input resources.

Population dynamics also impact the ability to catch each animal. Despite the present-day issues with fish sustainability, we assume that fish can multiply suitably fast to facilitate infinite foraging. This is designed to avoid the issue of permanent resource depletion, which would otherwise cause the system to collapse trivially (i.e., we are concerned with contingently existential threats that are avoidable by voluntary social construction, rather then necessarily existential threats that might require more draconian methods). Conversely, we both limit and impose a rate of reproduction on the deer population. This has two effects: first, it does allow the deer to be hunted into extinction, dissuading continual foraging as a possible strategy, and second, it gives salience and comparison to forecasting, where signaling must be used to communicate to other players that the resource is nearing depletion and benefit accrues to those players whose forecasts are most accurate.

After the set of games is complete, the total utility gained from all games (Equations 1 and 2) is summed and redistributed proportionally to the total resources invested in the hunt (Equation 3).

The dilemma in this case is a matter of risk. Foraging fish allows for a guaranteed low return of resources, whereas foraging for deer gives a lower chance at a higher payoff. For this reason, the strategy for foraging fish is *risk dominant*, whereas the strategy for foraging deer is *payoff dominant*. In a scenario where the survivability of the collective is paramount, the stag hunt game becomes a problem of "risky coordination." With few players coordinating for a high return, due to either a small number of players or many players acting selfishly, survivability is hindered, as the weaker players will be eliminated.

This leads to player utility being a function of foraging choice $c_i \in \{d,f\}$, numerical payoff $H_{c1,c2}$, the probability of a successful hunt $p(c_i)$, the number of participants N, the resources contributed by each player to the forage r_i, and the utility tier of the fishing expedition $UT(r_i)$.

For this reason, we define the players' utility function as follows. For N players $i \in \{1,\ldots,N\}$ forming $N/2$ iterated stag hunt games $G \in \{1,\ldots,N/2\}$ on each iterated round $t \in \{1,\ldots,\infty\}$, we define the payoff ($G'$) of game G played by player i as

$$G'_i = H_{c_i,c_j} * p(c_i) * n(c_i, r_i) \tag{1}$$

$$n(c_i, r_i) = \begin{cases} UT(r_i) & \text{if } c_i = f \\ 1 & \text{if } c_i = d \end{cases} \tag{2}$$

where $n(c_i, r_i)$ represents the "number" of animals foraged, showing that only one deer may be foraged per game, however, multiple fish can be foraged, depending on the utility tier (UT) discussed in section 4.5.3. Furthermore, c_j represents the foraging choice for the second player playing the stag hunt game. This provides the utility for each player as

$$u_i^{sh} = \frac{r_i}{\sum_{j=1}^{N} r_j} * \sum_{j=1}^{N} G'_j \tag{3}$$

illustrating that the total utility a player receives from a foraging session is proportional to the resources that they input. This results in two "temporary pools" being formed, one each for the stag hunters and fishermen, which are used to distribute resources according to Equation 3. This is not to be confused with the common pool, which is permanently present for all players and is designed as a reservoir for surplus resources.

2.3.2 Common-Pool Resource Management

A CPR management problem is a type of iterated collective action situation that investigates the issues concerning provision to and appropriation from a shared resource (Ostrom, 1990).

Regarding access, we consider two classifications: exclusivity, which concerns the ability to exclude individuals from the benefits of the resource, and subtractability, which determines the extent to which the benefits consumed by one individual subtract from the benefits available to others. The CPR in this article is classified as subtractable, as appropriation from the common pool is disadvantageous to the good of the collective, but not exclusive, as all players may make appropriations, irrespective of permission.

A side effect of a common pool is that each player trying to satisfy themselves maximally may conflict with the collective goal of sustainability of both the resource and the players. Therefore the players need to be *satisficed* (satisfied minimally) at least to some degree (over time); however, attempting to satisfy all the players maximally may deplete the resource.

The dilemma hence becomes a matter of personal security, with players having to evaluate the short- and long-term impacts of either saving resources to maximize a private pool or contributing to a global pot. Whereas the former is the rational short-term action for a reasonable, self-interested player, the latter creates the "safety net" of an accessible resource, should the players find themselves with a depleted private pool at some point in the future. This yields that the pool is sufficient to satisfice needs if the players manage to cooperate; however, there is simply not enough if they fail to do so.

This provides players with a set of possible strategies for playing the game. Individual utility of a player is maximized by withholding provision of their own resources on the assumption that all other players contribute fully (the Nash equilibrium); but a player can also demand the maximum appropriation irrespective of their actual need and can also cheat on the appropriation of resources. This prompts the need for conventional (institutional) rules to regulate behavior but also observational, sanctioning, and dispute-resolution mechanisms to check for and punish noncompliance (cf. Pitt, 2021, chapter 5).

For this system, we take N players $i \in \{1, \ldots, N\}$ that perform the following actions in each iterated round $t \in \{1, \ldots, \infty\}$: (a) makes a provision of resources due to sanctioning, s_i; (b) makes a provision of resources through tax, t_i; (c) makes a demand for resources, d_i; (d) receives an allocation of resources, a_i; (e) makes an appropriation of resources, a'_i; and (f) receives a salary for roles in government g_i. The total resources accrued at the end of a round is hence the utility of a player and is given by

$$u_i^{cpr} = a'_i + g_i - (s_i + t_i) \tag{4}$$

2.3.3 Collective Risk Dilemma

By creating a common pool of resources, there is a facility to mitigate a disaster (a catastrophic event that, if unmitigated, has the power to inflict great damage across the multiagent system), so long as a threshold is reached. This converts a simple Linear Public Goods (LPG) game into a twofold problem known as the *Collective Risk Dilemma* (CRD).

This dilemma, a subset of games known as *threshold public goods games* (Cadsby & Maynes, 1999), provides N players with T rounds to reach a fixed threshold. In the event of reaching this threshold, the individual damages incurred in the event of disaster can be mitigated to reduce the effect, providing a collective benefit in the form of an increased likelihood of survival. Ultimately, the CRD aims to conflict the individual, short-term benefit of resource retention with the collective, long-term benefit of disaster mitigation. Rationally, self-interested behavior dissuades contribution to a common pool; however, this would be detrimental to the collective, as having all members contribute is the optimal strategy when the resource loss incurred from unsuccessful mitigation outweighs any short-term benefit.

In this article, we increase the difficulty of the CRD by varying the uncertainty of the system—we consider the need for external self-organization when first the threshold value that needs to be reached and second the knowledge of the periodicity of disaster are visible or hidden. The latter gives rise to the notion of timing uncertainty stipulated by Domingos et al. (2020), who identify that in real-world scenarios, the quantity needed to contribute and the time when it must be achieved are often not certain and instead based on predictions.

We model the expected utility for each player $i \in \{1, \ldots, N\}$ as dependent on the total resources in the common pool R, the threshold needed to mitigate disaster TH, the geographical position of the player p_i, and the magnitude of the disaster M. The resulting utility is hence

$$u_i^{crd} = \begin{cases} 0 & \text{if } R \geq \mathit{TH} \\ -U(p_i, M, R) & \text{otherwise} \end{cases} \tag{5}$$

where the utility mapping $U(p_i, M, R)$ is specified in section 4.1.

2.3.4 Total Utility Gain

Through engaging in the three games detailed earlier, each player $i \in \{1, \ldots, N\}$ gains a total utility per turn u_i^*, as computed by Equation 6:

$$u_i^* = u_i^{sh} + u_i^{cpr} + u_i^{crd} \tag{6}$$

Note that each of these three operational-choice resource-competition games has a Nash equilibrium solution that operates to the detriment of the others. This solution concept in one game has an adverse knock-on effect on another; therefore the Nash equilibrium is no guarantor of cooperative survival in this context. Note that removing any game would remove the element of social construction that impacts decision-making in the others. Simply offering players an endowment would remove the need for social coordination, and removing any of the two dilemmas would trivialize the problem: The lack of a CRD means that all resources exist in a closed system, which simplifies the resource management problem. Instead, we introduce self-organizing mechanisms that effectively create three social-choice political-competition games that socially construct resources (institutions, social capital, etc.) that have side effects (change the constraints) on the operational-choice games (but that is not to say that the social-choice games cannot be "gamed" either, e.g., by autocratic takeover of governance).

2.4 Self-Organizing Mechanisms

The games introduced in the previous section introduce various social dilemmas. For example, the stag hunt game presents players with a dilemma of risk minimization versus reward optimization. The CPR situation is essentially a problem of short-term response escalation leading to long-term ruination, that is, the (supposedly inevitable) tragedy of the commons. Both the CPR and CRD games present an issue of free riding: The optimal strategy is not to contribute and to rely on everyone else providing enough, which works fine for one practicing it—provided there is only one. If everyone adopts this strategy, there is disaster for all. To overcome these problems or avoid socially suboptimal outcomes, various mechanisms have emerged based on self-organization through the social construction (Berger & Luckmann, 1966) of rules, relations, and reputations.

2.4.1 Governance

The social construction of rules to help resolve a social dilemma or collective action problem has been a key tool to help communities solve political problems, such as those identified by Plato in

The Republic, for example, how to prepare the "best suited" to rule for the common good and to prevent those ill suited to the task from occupying political office.

Subsequently, Aristotle wrote of developing a paradigm of constitutional self-government, combining both "aristocratic" and "democratic" elements (Durant, 2006). This approach to governance gave rise to the classical Athenian political regime of democracy, in which the citizens can be characterized as *collective rulers*, delegating governmental duties according to a system of votes, including sortition (the selection of public officials by random sampling from a larger population). In classical Athenian terms, this represented ruling, as citizens, and being ruled over, by other citizens, in turns.

However, the ongoing development of the principles and practice of democratic self-governance through history can be viewed as a gradual evolution along a continuous spectrum into a better form, or evolution as a series of punctuated equilibrium states, akin to regular "software updates" (Manville & Ober, 2019). Either way, though, the social construction of sets of rules—that is, the mutual agreement of conventional rules which serve to regulate or constrain behavior—is not the sole preserve of city-states or national governments.

In particular, Ostrom's extensive fieldwork collected rigorous empirical evidence to demonstrate how communities, throughout history and geography, could develop a long-lasting solution to a collective-action problem involving sustainable CPR management based on the formation of self-governing institutions (Ostrom, 1990). Moreover, these communities were able to coordinate their activities with others to solve such problems at scale in a system of systems (see also Lansing & Kremer, 1993). The common features of Ostrom's self-governing institutions can be expressed as design principles, and the principles themselves have been reexpressed algorithmically for use in electronic institutions (Pitt, 2021).

The need for, and benefit of, some form of governance as a mechanism for self-organization is further demonstrated by the codification of *rules of order* for deliberative assemblies (Robert et al., 2000). These rules are applicable from parish councils to national parliaments, and this work provides a standard handbook documenting best practice in all such situations. However, many other situations are also covered by socially constructed rules, from contracts of employment through terms and conditions of service up to international treaties and the rule of law.

2.4.2 Trading

A second form of self-organization, providing the wherewithal to relax the constraints of these games, is the social construction of conceptual relations, for example, though interactions that create *externalities*. In economics, generally speaking, externalities are benefits that accrue to a third party as a result of interaction between two other parties; however, successful interactions can also provide externalities that benefit both parties, for example, creating trust relationships (which can also be reported to the benefit of other parties). These externalities have been called *social capital* (Putnam, 2000), but because of the connotations associated with that term, the term we use here is *conceptual resources* (i.e., these are abstract, socially constructed resources in the "minds" of the agents, not physical resources like fish or deer).

For example, Malinowksi's (1920) anthropological research studying the Kula ring exchange system in the Trobriand Islands demonstrated how trading established cohesive bonds between otherwise disparate groups. The Kula ring spanned 18 island nations across the Massim archipelago of Papua New Guinea and involved thousands of participants.[1] All Kula valuables are handmade trinkets that are traded solely for the purpose of currying favor in the archipelago and boosting an island's social prestige. So, an islander would make an arduous journey by canoe to present a gift to a (higher-ranked) person on another island. However, it was considered a social faux pas to retain a Kula gift, and so the second person would make an arduous journey by canoe to present the gift to yet another person on a third island. This would be repeated, forming a ring around the islands. The

1 In passing, it is Malinowski's (1920) description of the Kula ring that inspired the archipelago as the basis for the scenario used in this article.

system of continuous trade resulted in a form of "gift economy" whose crucial feature was to create reciprocal relationships so that in times of hardship, one island would be incentivized to support another.

In this way, trading creates a deeper context that essentially affects motivations, decisions, and behaviors in social dilemmas. The giving of gifts, of different value and frequency, serves to create an *economy of esteem* (Brennan & Pettit, 2005). The crucial feature of "esteem" is that it cannot be traded—it can only be produced—but it can have a significant effect in free markets and other social dilemmas. For example, it has been shown in an iterated time-slot scheduling system that exchanges based on the production and calling in of "favors" enabled a completely decentralized system with no central authority to approximate the optimal solution. Any differences were offset by the lower cost of the increases in communication compared to the cost of computing that optimal solution (Pitt et al., 2014). In fact, many problems of collective action and knowledge management rely on lowering transaction costs, by grafting the prosocial behaviors (e.g., giving gifts or doing favors) onto actions that people would have done anyway (Ober, 2017).

2.4.3 Forecasting

Instances of forecasting or divination, more or less scientifically, to support or guide decision-making can also be seen throughout history. For example, in classical Greece, oracles were supposedly gifted people who were thought to be able to channel inspiration from the deities to generate prophetic predictions of the future. In ancient Delphi, Greece, the Pythia spoke for the oracles of the god of the sun and light, Apollo. She responded to the questions of people from all walks of life, whether citizens or kings, on diverse issues ranging from political impact and war to laws and personal issues (Broad, 2007). Overall, the Pythia was regarded as the highest civil and religious authority, influencing both the government's and individuals' decisions on all important topics.

However, soothsayers, shamans, witch doctors, priests, and so on, have all claimed some form of divine insight for predicting the future. For example, during the 17th century, plague doctors began to prevail as a means of forecasting treatment for the Great Plague of London. They were hired by cities to treat plagued patients from all backgrounds, though they rarely succeeded, serving instead to record death tolls and the number of infections (Byrne, 2006). Again, these plague doctors would serve as the local authority on medical treatment, helping influence the government on which methods were effective at the time. Although such activity took place before the germ theory of disease was properly articulated, it can be seen as the forerunner to statistical epidemiology, which uses rather more informed theories and techniques to help formulate public health policy.

Feedback is a critical feature of forecasting, and in economics, scoring rules have been proposed that give credit (or otherwise) to forecasters for the accuracy (or lack of it) of their predictions (Harrison et al., 2017). One of the particular problems with forecasting, though, is that predictions can affect behaviors that give rise to different outcomes. For example, a weather forecast for *rain* may influence someone to carry an umbrella; it would be inappropriate to accuse the forecaster of being incompetent because they did not get wet.

2.5 Summary: The Challenge of Cooperative Survival

In this context, the challenge to ALife becomes not just a matter of life but a matter of life and death: Which combination of what form of these self-organizing mechanisms "solves," optimally or otherwise, the cooperative survival game under what dimensions of scale, uncertainty, and constraints of the subgame. In the next section, we design a self-organizing multiagent simulator to address this challenge.

3 System Design

This section details the codification of the overall "platform" for this multiagent system. As such, we discuss the sequence of actions that drive this simulator in section 3.1 by presenting it as a

state machine, followed by a discussion of the agent architecture in section 3.2 and how agents may interact with the different self-organizing mechanisms. We conclude this section by discussing the representation of rules for the governance mechanism in section 3.3.

3.1 Simulator Design

We implement an iterative simulation in which all islands in the archipelago play up to three political metagames each turn. The simulation is structured as series of seasons and turns, which conclude once all the islands "die." The objective for the islands is to survive for as many turns and seasons as possible. We set up our simulation such that agents will play the games for 50 days, with disasters affecting the archipelago every 5 days. This yields 10 seasons of play.

A *turn* is defined as a series of exchanges between agents in which islands receive resource updates, attend Inter-Island Organization meetings (see later), and interact with one another and the game state through *actions*. A turn can be broken down as follows:

1. Resource updates are given for each island and on the global game state.
2. The Inter-Island Governmental Organization (IIGO) decides rule changes, elections, and sanctions.
3. The Inter-Island Forecasting Organization (IIFO) provides a forum for information exchange to mitigate both short- and long-term risk dilemmas.
4. The Inter-Island Trade Organization (IITO) facilitates gift exchanges and allows agents to communicate to make deals between one another without organization supervision.
5. Islands submit decisions on their actions to the server to formally end the turn.
6. A check is made to see if a disaster occurs this turn.
7. The server processes actions and updates game and island states:
   a. A cost of living is subtracted from an island's pool before the next term. This is the simulation-level equivalent to using resources to stay alive (e.g., food consumed). These resources are permanently consumed and do *not* go into the common pool. Note that this is *not* the same as the tax.
   b. Check if the game is over.
   c. Check if any islands are *critical* (i.e., below the threshold).
   d. Check if any islands are *dead*.

Furthermore, a *season* is defined as a series of turns and concludes with a disaster. Seasons formalize the flow of the game and provide a method to track the number of disasters the islands survive. We illustrate our formal definition of a turn in Figure 3.

3.2 Agent Design

All agents $i \in A$ are implemented as a data structure such that all agents contain sufficient parameterization to participate in the various metagames $g \in G$, resulting in $I = < A, G >$.

Each agent must have the capacity to (a) identify other agents and (b) interact with the self-organizing mechanisms of *forecasting*, *gifting*, and *governance*. Agents are ascribed a unique ID, represented by an integer from 1 to the total number of agents, that remains constant throughout the entire simulator.

For interfacing with the IIGO, a global rule cache stores the rules that are available to the IIGO, and a local rule cache stores those that are in play. The representation for these rules are later discussed in section 3.3. Agents also maintain knowledge of the agents that hold power in the IIGO, as this process is not anonymous.

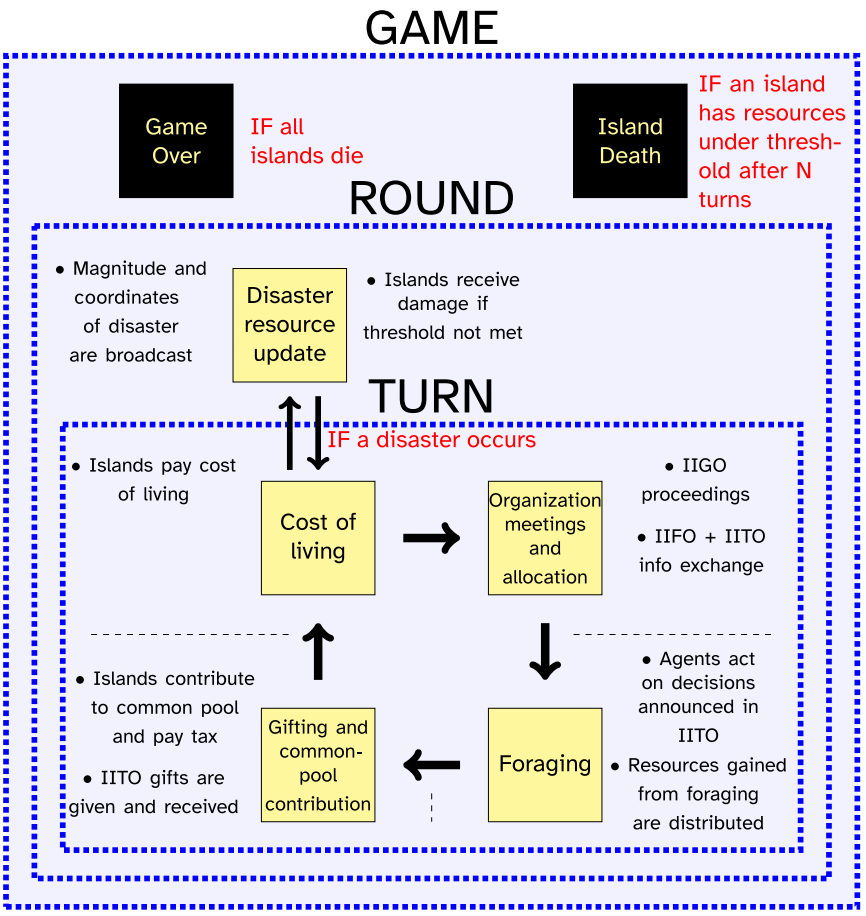


Figure 3. Visualization of a turn. IITO = Inter-Island Trade Organization; IIGO = Inter-Island Governmental Organization; IIFO = Inter-Island Forecasting Organization.

Furthermore, for interaction with the IIFO, agents make a *disasterPrediction*, as shown in Table 1. This stores predictions concerning the position, magnitude, and periodicity of disaster, as well as the confidence that the agent has in its prediction.

Finally, to engage in the IITO, agents must first engage in opinion formation. Table 2 demonstrates the variables that are required for this: An agent's decision to report resources must be tracked to inform their trustworthiness, as shown in the variable *AgentReportedResources*. This maps each agent's ID to a Boolean corresponding to whether an agent has reported their resources on that turn. In addition, the gifts that have been received from other agents affect the quality of gifts that an agent will, in turn, offer. This is recorded in the *ReceivedOffer* variable, which again tracks the mapping of agent IDs to numerical gifts.

Furthermore, throughout the simulator, the opinions that agents have of one another fluctuate based on gifting and lawful behavior. This is tracked in the variable *AgentOpinion*, where the agent ID is mapped to a numerical "score," defined on $[-maxOpinion, maxOpinion]$, with larger, positive values corresponding to a higher opinion (parameterization is discussed in section 4.2.2). The range of possible opinion values is constrained such that it is not possible for successive negative interactions to cause opinion to "spiral out of control." This effectively gives agents the opportunity for "redemption" (or, contrarily, it means that they cannot act kindly in the early stages of the game in order to act tyranically later on, while still reaping the rewards of a high trust value).

Table 1. Parameters held in the *disasterPrediction* data structure.

Parameter	Range
Coordinate X	$N \in [0, 10]$
Coordinate Y	$N \in [0, 10]$
Magnitude	$N \in (0,]$
Confidence	$N \in (0, 1]$
Threshold	N
Period	N

Table 2. Parameters held in the *Agent* data structure.

Parameter	Range
AgentReportedResources	map{*A, Bool*}
ReceivedOffer	map{*A, N*}
AgentOpinion	map{*A, N*}

3.3 Rule Representation

Rule generation represents a large portion of the function of the governmental organization. For this simulation, we represent our rules as a set of vectors such that the principles of linear algebra can be applied. The premise of this choice was such that, if a rule could be represented by matrices, then agents could look up every element of the matrix with minimal complexity.

Following is an example of this matrix-based rule representation:

Rule 1. *If you are expected to pay x amount of tax, the amount of tax you pay must be x.*

If our goal is simply to ensure that an agent pays the expected amount of tax, a basic mathematical operation we can perform is to subtract the actual tax paid from the expected. If we obtain 0 from this subtraction, we can conclude that this agent has adhered to this rule. Formally, letting x and y represent the expected and the actual paid amount of tax, respectively, we calculate the following:

$$y - x = 0 \tag{7}$$

To turn Equation 7 into a matrix calculation, we introduce some trivial coefficients:

$$-1 * x + 1 * y = 0 \tag{8}$$

Now, we can write Equation 8 as a matrix calculation:

$$\begin{bmatrix} -1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 0 \tag{9}$$

Following is a more complicated example:

Rule 2. *If you are expected to pay x amount of tax, the amount of tax you pay must be at least x.*

Table 3. Auxiliary codes for rule encoding.

Auxiliary code	Meaning
0	= 0
1	< 0
2	≥ 0
3	!= 0
4	real

We can rewrite the matrix equation as an inequality:

$$\begin{bmatrix} -1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \geq 0 \tag{10}$$

The addition of the "at least" (i.e., ≥) in Equation 10 adds complexity to the proposed matrix scheme. For this reason, it is not possible simply to check for equality, and although all calculations are *compared* to 0, further encoding is needed to see how the comparison is made. To achieve this, we incorporate an *auxiliary* vector containing a list of comparison codes, as seen in Table 3. However, with this approach, problems are still encountered with a rule like Rule 3.

Rule 3. *If you are expected to pay x amount of tax, the amount of tax you pay must be at least $x + 5$.*

To capture such cases of rules, we introduce a constant to the input list, as in Equation 11:

$$\begin{bmatrix} -1 & 1 & -5 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \geq 0 \tag{11}$$

This now allows the encoding to capture any linear equation or inequality with any constant shift. It is important to note that the dimensions of this input vector can be arbitrarily large, as for each additional constraint a rule has (and a variable needed to fill), we expand the input dimensions of our vector.

4 Self-Organizing Mechanisms and Games

This section discusses the methods for codifying the aforementioned self-organizing mechanisms and games by specifying the algorithms.

4.1 Disaster

All agents are distributed across a 2-D plane ranging from (0, 0) to (10, 10). For simplicity, the common pool is interpreted to have a presence outside the realm of the games such that the damage dealt to the common pool is independent of the location of the disaster. Naturally, a well-maintained common pool will allow for the maximum mitigation of damage from the disaster (the resources held by the agents will minimally deplete), incentivizing the agents to self-organize for cooperative survival.

The disaster generates uniformly across the 2-D plane as a point and damages the agents according to two "effects": The *soloEffect* is calculated according to the magnitude of the disaster divided

```
totalEffect ← ∑ soloEffect
if currentCP ≥ thresholdCP then
    impact ← 0.5 ∗ totalEffect
else
    impact ← totalEffect
end if
for all agents ∈ aliveAgents do
    if currentCP ≥ impact then
        damageFelt ← 0
    else
        remainingDamage ← impact − currentCP
        damageFelt ← remainingDamage ∗ propEffect
    end if
    agentResources ← agentResources − damageFelt
end for
currentCP ← currentCP − impact
```

Figure 4. Algorithm 1: resource deduction due to disaster.

by the squared distance to each agent, which is subsequently converted to a *propEffect* by taking the ratio of the *soloEffect* to the total damage felt across all agents. Disaster mitigation is achieved depending on the common-pool threshold, *thresholdCP*, and the resources available in the common pool, *currentCP* (TH and *R*, respectively, in Equation 5). The resource deduction to the agents, *agentResources*, and common pool is then calculated according to Algorithm 1 (Figure 4).

In this algorithm, if the resources in the common pool are above threshold, the impact is halved for all agents. This incentivizes filling the common pool to at least the baseline, as the total damage can be somewhat mitigated. Following this, the resources in the common pool are compared with the aforementioned impact, where if the resource availability exceeds the impact, the damage taken is 0. This incentivizes *overfilling* the pool. Ultimately, this algorithm rewards both a minimal filling of the pool and an overfilling, thus encouraging self-organization.

4.2 Trading

Trading serves as the main self-organizing mechanism for facilitating opinion formation over the social network, which in turn influences the electoral process for introducing politics to the multi-agent system. We break down the process of gifting into two main stages: sending gifts and receiving gifts.

4.2.1 Requesting Gifts

Each client has the capacity to make a gift request. We model the gifting system as a set of requests made by the agents, based on their private pool of resources. If an agent is below the critical

threshold, they will ask all other agents for a gift equal to three times the cost of living. Alternatively, an agent will have a 40% chance of asking each other agent for a gift equal to the following:

$$ReceivedOffer[agent] = 2 * (AgentOpinion[agent] + costOfLiving) \tag{12}$$

4.2.2 Receiving Gifts

Gift response sessions are handled such that all requests are seen at the same time, by passing a list of gift requests to each agent. Upon receiving this list, the agent will sort the requests made by descending opinion, allowing for agents to prioritize their favored other agents. After sorting, one of three possible responses is made:

1. An agent will respond with a gift of 0 in three cases: either if (a) the private resources held by the agent are less than the *anxiety threshold* (250, for this article), (b) the *current (private) resources* held by the agent are less than the request being made, or (c) the opinion held of the agent making the request is less than 0.
2. If not responding with a gift of 0, an agent will respond with a gift equal to the request in two cases: (a) the opinion held of the agent making the request is equal to the *maxOpinion* (30, in this case) or (b) the agent making the request has resources below the critical level.
3. If neither of these two cases passes, the default gift request is handled as follows:

$$response[agent] = min\left(ReceivedOffer[agent], \frac{currentResources^2}{100 * anxietyThreshold}\right) \tag{13}$$

4.2.3 Opinion Formation

It is the process of receiving and proposing gifts that permits agents to develop opinions of the other agents and, by extension, form a social network in the simulator. The magnitude of change in the opinion held of an agent during trade is based on the response received to a proposed gift request.

If the response to the request is nonzero but less than the requested amount, the opinion held of the responding agent increases by a single point. If instead, however, the response is greater than the request, the opinion held of the responding agent increases by 20% of *maxOpinion*. Furthermore, if the agent *proposing* the gift request is in a critical state (that is, has resources below the critical value) and receives a nonzero response, the opinion held of the responding agent increases by 50% of the *maxOpinion*. In any case, additional generosity is encouraged, as it facilitates a faster development of positive opinion. Finally, if the received amount is zero, then the opinion held of the responding agent *decreases* by a single point. If all of these cases fail, the default strategy is not to update the opinion at all.

4.3 Governance

Defining social choice political metagames to regulate behavior in operational-choice resource-competition games requires collective self-governance, and this self-organizing mechanism requires conventional rules (institutions; Ostrom, 1990) and methods to select, revise, apply, revoke, and enforce those rules. Given that agents are assumed to be essentially equal (in territory, mineral resources, person power, technological development, etc.), we follow Ober's Demopolis thought experiment (Ober, 2017; Pitt & Ober, 2018).

There are two points to note about governance: structure and representation. First, regarding structure, following the "standard" (and typically ascribed to Montesquieu) blueprint for the

separation of powers and checks on the use of those powers, government in the archipelago is represented by three distinct institutions: the *legislative institution*, which makes rules; the *executive institution*, which applies rules; and the *judicial institution*, which interprets rules.

Second, regarding the representation of those separated powers, we follow the standard representation in norm-governed multiagent systems with institutional facts (Artikis et al., 2009) to distinguish between institutionalized power (as opposed to physical actions) (Jones & Sergot, 1993), permission, and obligation.

Institutionalized power is the performance of a designated act (typically but not necessarily a speech act) by an assigned agent occupying a specific role, which "counts as" an institutional fact being (mutually agreed) to be true. Then, institutionalized power in the three government institutions is assigned to those agents occupying roles of *Speaker* in the legislature, *President* in the executive, and *Judge* in the judiciary.

By performing designated actions, some of which are permitted (sometimes, but not always, power implies permission) and some of which are obliged, an agent occupying a role sees to it that certain institutional facts are true, as discussed in the subsequent sections. Note that as actions are sequential, we can use the current state to compute what power, permissions, and obligations apply to an agent in that state, without using the *Event Calculus*, for example (Artikis et al., 2015).

4.3.1 Roles and Powers

To help distinguish between institutionalized power and physical capability and between (institutional or physical) power and permission, consider the role of *Judge*. Only the agent occupying the role of *Judge* can declare that an audit of another agent is to be carried out: Any other agent making such a declaration is merely making noise. The *Judge* is permitted to conduct an audit, which is a physical action, but only if an audit has been declared. Only the agent in the role of *Judge* can declare the result of an audit; moreover, an agent is not permitted to declare the result of an audit of itself (reflecting the principle of *nemo judex in causa sua*, that no one should be a judge in their own case).

In addition, the agent in the role of *Judge* is empowered to impose sanctions, but is only permitted to impose sanctions if the audit has revealed a breach of the rules—in this case, power does not imply permission. Note that if the *Judge* agent does impose a sanction without permission, the institutional fact that there is a sanction is still true; it is the *Judge* agent that has violated the normative rules. For this reason, there should be an appeals procedure (Ostrom, 1990), but for simplicity, in this system, agents are effectively regimented (Jones & Sergot, 1993) and cannot do what is not permitted.

The agent occupying the role of *President* is empowered primarily to do three things: to set the level of taxation, to declare the allocation of resources, and to set the agenda for the legislature. Note that only the agent occupying this role can make declarations that count as determining the tax level, the resource allocation, or the agenda: Any other agent making such actions is again making noise.

Similarly, the powers of the agent occupying the role of *Speaker* are primarily associated with the conduct of elections and the declaration of results. Only the agent in the role of *Speaker* is empowered to declare the result of a vote, and so only then does a rule become active, revised, or revoked. Note that there are also obligations associated with this role, most importantly, that the *Speaker* is obliged to select at least one rule for voting if rules have been selected by the *President* and that the *Speaker* is obliged to declare results according to the way votes were cast (cf. Pitt et al., 2006).

4.3.2 Voting

Proceedings in the governmental organization are based on a simple voting system, where all agents will compile a ballot for the topic in hand to be sent to the *Speaker* for broadcasting. Voting occurs in two instances: When choosing to implement a new rule and when appointing the new roles in government. All agents are programmed to vote in favor of a proposed rule. When voting on elections, however, the agents implement a *Borda count* method, where N agents are ranked from

highest opinion to lowest opinion. The Borda score for each agent a is then computed by Equation 14, where b represents the number of ballots:

$$\sum_{i=1}^{b} N - \text{rank}(a, v_i) + 1 \tag{14}$$

where $\text{rank}(a, v_i)$ is the position (rank) of agent a in the ballot v_i of agent i.

4.3.3 Taxation

Governance serves as a self-organizing method for replenishment of the common pool. We take the agent currently elected as *President* for defining *taxation*, where a percentage of all agents' resources are taken each turn. Depending on the availability of the forecasting organization, this taxation can be either informed (based on previous assessment of the common pool) or random (assuming that the forecasting organization is deactivated).

Uncertainty is introduced to the system in two ways—through stochastic period and magnitude: Disasters can occur regularly, with a known interval period, or randomly, obeying a geometric distribution such that the expected period is the fixed period of the regular case and *variably visible common pools*—agents may or may not have knowledge of the common pool's threshold. The strategy for taxation is hence dependent on the uncertainty of the system and a set of tuning parameters that define the numerical contribution needed with respect to the games.

The ability to use the forecasting organization is not always available, hence the elected *President* of the IIGO makes a prediction for the period of disaster and the common-pool threshold on both the first turn of the game and any subsequent turns when a disaster occurs. This guess is randomized from [2, 10] for the disaster periodicity and from [200, 1000] for the threshold. Naturally, this is an inferior strategy to using the forecasting organization, as both of these guesses run a severe risk of overestimation, meaning that the common pool will either not be filled in time or heavily overcontributed to, resulting in wasted utility.

The general principle for taxation is to predict either of these quantities on each turn, CPT'_t and T'_t, whether they be visible, estimated by the *President*, or estimated by the forecasting organization, giving the "quality" of taxation as *known* > *forecast* > *President*.

Simply, if a quantity is *known*, then its prediction is equal to the ground truth. If a quantity is hidden, then its prediction will be equal to the guess made by the forecasting organization in section 4.4.1, if active, or else the *President*.

The total required contribution per agent per turn $c_{i,t}$ with $i \in \{1, \ldots, N\}$ is ultimately defined in Equation 15, with N_t representing the number of *alive* agents and CP_t the resources in the common pool on turn t. The *overfillRate* represents the "multiplier" applied to the common-pool threshold to request agents to contribute resources surplus to the minimum threshold,

$$c_{i,t} = \frac{(CPT'_t * \mathit{overfillRate}) - CP_t}{N_t * T'_t} \tag{15}$$

to represent that the taxation per turn should be split evenly across all alive agents and all turns leading up to the next disaster.

4.3.4 Sanctioning

Sanctions are interpreted as a *fine* imposed on any agents who break a rule during the simulation run time.

Sanctions are assigned based on five increasing tiers of severity 1–5. A sanction tier is allocated based on the running total of rules broken by an agent, with the thresholds defined as 1, 5, 10, 20, and 30, respectively.

The sanction tier defines the severity of fine imposed on the agent. This payment must be made directly to the common pool, as avoidance will result in further sanctions being applied. The sanction requirements for the tiers increase with severity and are numerically represented by a base cost of 10 resources, plus a fraction of the current resources. Increasing from Tier 1 to Tier 5, the additional fractions are 0%, 20%, 30%, 50%, and 80%, respectively.

4.3.5 Order of Proceedings

The overall process for governance follows a linear path. Upon initiating the governmental organization at the start of a turn, events proceed such that the *Judge* will evaluate previous performances, the *President* will take appropriate action to deal with the *Judge's* findings, and the *Speaker* will announce the actions carried out by the *President*.

The governmental organization has a strict order of proceedings, as follows, with all branches working in tandem.

4.3.5.1 Initial Proceedings At the start of each turn, the payment made to each governmental branch from the common pool is increased if agents have voted to increase the respective budget. Following this, any agent occupying a role of power has the counter tracking their total duration spent in power incremented by 1.

4.3.5.2 Judicial Branch The *Judge* begins by selecting an agent in the system and inspecting their decision history (such as tax contributions) to evaluate if the agent has complied with the rules throughout the simulation. Following this, the *Judge* is able to define the severity of the sanction needed and to apply it accordingly. Importantly, the *Judge* will not take their own previous actions into consideration, making them exempt from sanctioning.

4.3.5.3 Executive Branch The agent occupying this branch begins by compiling a report concerning what each agent (a) possesses and (b) *claims* to possess in its private pool. This is then passed to the *Judge* to reevaluate if an agent should be pardoned. Importantly, if an agent refuses to report their current resources, the shared opinion of this agent decreases by one stage (section 4.2.3), under the assumption that if an agent refuses to report their resources, the agent is unlikely to pay tax as well. Following from this, the *President* broadcasts the amount of tax needed to be paid by each agent, according to Equation 15. This serves as the key mechanic for developing the resources in the common pool.

The *President* also facilitates allocations from the common pool, as well as taxation. Resource redistribution from the common pool occurs if a request is made, where agents will ask for an allocation if they have 300 resources or fewer, making a request for an allocation of 50 resources. The allocation is calculated based on the total request made by all agents. If the total request is less than 75% of the common pool, the individual requests are granted without question. If this is not the case, the allocation to agent i on turn t is granted according to Equation 16:

$$allocation_{i,t} = request_i * CP_t * \frac{3}{4 * totalRequest_t} \tag{16}$$

Finally, all agents in the system are polled for a new rule that they wish to be implemented in government. The *President* selects at random a rule from this list to be supplied to the *Speaker* for broadcasting to all agents.

4.3.5.4 Legislative Branch Concluding the governmental proceedings is the legislative branch. Upon broadcasting the newly proposed rule, all agents will cast their votes to offer a decision to accept or decline the rule. This decision will hence inform if the new rule is added to the

list of active rules. All agents will then vote on who they want to act in government, according to Equation 14. This result is then announced by the *Speaker*.

4.4 Forecasting

4.4.1 Disaster Mitigation

When the common-pool threshold is invisible, self-organization through forecasting serves as a means of helping counteract the increased uncertainty imposed on the multiagent system. When forecasting is active, agents will make regular guesses at the value of the common-pool threshold and the period of disaster based on the damage that they have received. If the damage taken is too high (>200), agents will increase their estimate of the common-pool threshold, hoping to further mitigate the damage taken. Conversely, if the agents seem to be taking little to no damage from the disaster, they will decrease their estimate of the common pool to reduce the resources lost from taxation. Both of the updates to this prediction are randomized in the range [50, 100]. The disaster period is estimated by averaging the periods of all previous disasters.

In the event that a disaster is yet to have occurred, the predictions are initialized with the common-pool estimation in the range [0, 1000] and the disaster period equal to the first turn on which it occurs.

Following a forecasting session, the final predictions made are averaged evenly across all agents, offering a "wisdom of the crowds" approach (Surowiecki, 2004), as an average of the guesses of all agents is more powerful than any individual guess.

4.4.2 Improving Foraging

Forecasting, furthermore, helps with the quality of foraging. In each forecasting session, agents selectively broadcast the previous foraging choice that they made to other agents. When this mechanism is active, information is given about the natures of the other agents, allowing inferences to be made about if the deer population is becoming over foraged. This ultimately has the effect of helping reduce the risk of hunting deer to extinction.

4.5 Stag Hunt and Foraging Returns

We codify the conventional stag hunt dilemma such that all agents have no a priori knowledge of the other agents' decisions, using only their interpretations of the population sizes shared by the *forecasting* organization. Three foraging types are conditionally implemented based on the environment:

randomForage. We introduce an equiprobable chance of either foraging a deer or choosing to fish. This forage type is chosen for the first five turns to build up knowledge of the randomized initial population or if forecasting is disabled.

desperateForage. In the instance that an agent has a critical level of resources, they constantly forage deer, hoping that another agent has made the same decision, hence looking for the highest possible payoff.

flipForage. Agents forage the opposite of what the mass did the previous turn and forage deer with an input amount inversely proportional to the sum of contributed resources. This foraging method is chosen as the default, assuming the criteria are not met for the previous two methods.

This foraging decision also helps define the *input resources*, which represents the number of resources an agent contributes to helping fund the expedition: In the instances of either choosing a *randomForage* or *flipForage*, agents will randomly contribute up to 10% of their current total resources, using a uniformly distributed random variable. When choosing a *desperateForage*, however, the agent makes a last-ditch attempt to generate resources, contributing every remaining resource they have.

4.5.1 Fish

The utility gained from an *individual* fish, F, caught from foraging is modeled on the following normal distribution:

$$F \sim \mathcal{N}(1.45,\ 0.1^2) \tag{17}$$

We assume that, given the sufficiently large presence of fish, the catch rate will always be 100%.

4.5.2 Deer

The utility gained from an *individual* deer, D, caught from foraging is modeled on the following exponential distribution:

$$D \sim \exp(0.5) + 1 \tag{18}$$

We also propose that deer are harder to catch than fish, instead having a catch rate of only 80%. Again, the returns of the deer are modeled in a similar way, proposing a larger output scalar to incentivize foraging for deer. We also add 1 to the average output of the exponential distribution to prevent an output of 0.

To model the population dynamics of the system, we impose a rate of reproduction that acts as a function of the existing deer population. Defining the population per turn p_t, the maximum population size $p_{\max}$, and the reproduction rate r, we arrive at the population update function

$$p_t = p_{t-1} + r * (p_{\max} - p_{t-1}) \tag{19}$$

using a maximum population size of 20 deer to limit the feasibility of foraging and a reproduction rate of 40%.

4.5.3 Utility Tier and Payoff

For both foraging types, it is possible to obtain more than one per expedition—this is henceforth referred to as the *utility tier* and is synonymous with the total number of animals foraged per expedition. The utility tier takes the number of resources put into the foraging expedition, along with a *decayFactor*. This *decayFactor* (0.95 for fish and 0.9 for deer) aims to replicate the idea that, after investing resources to reach the foraging location, subsequent animals would be easier to catch, as no further resources need to be invested for changing location. The maximum possible utility tier (and hence the *maxNumberPerHunt*) for this article is 10 for fish and 5 for deer.

With an *inputScalar* again keeping the resource investment commensurate with the cost of living, the utility tier is calculated using Algorithm 2 (Figure 5), which yields the step function seen in Figure 6.

The resources obtained through foraging are calculated differently, based on the species in question. The input resources for each agent are taken as a random fraction of up to 10% of the total resources. There are two separate distribution strategies: The *equal split* applies to fishing and is implemented such that the resources are equally distributed across the number of agents participating in fishing; conversely, the *proportional split*, applied to hunting deer, distributes resources proportionally to the amount contributed to the hunt. On the basis of the previous sections, we generate the payoff matrix in Figure 2, with the rewards for foraging as either a deer (D) or a fish (F). Naturally, there is incentive for both agents to collaborate in the deer hunt so as to reap the maximal rewards.

4.6 Scaling

We note that the utilities gained from each game are kept within the same order of magnitude as each other. This ensures that no game is weighted above any other and that all have an equal

```
totalInput ← ∑_aliveAgents input
sum ← 0
n ← 0
while n < maxNumberPerHunt do
    sum ← sum + decayFactor^n ∗ inputScalar
    if totalInput < sum then
        return n
    end if
    n ← n + 1
end while
return maxNumberPerHunt
```

Figure 5. Algorithm 2: utility tier.

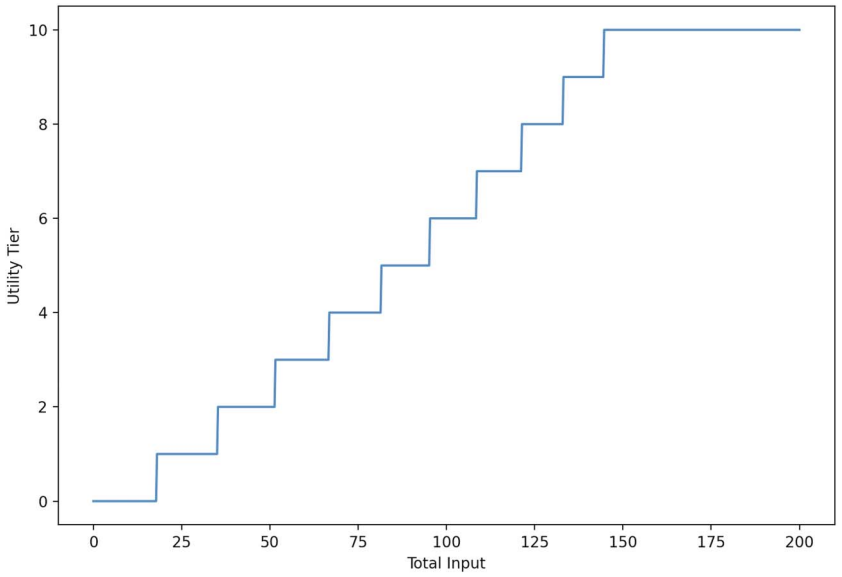


Figure 6. Utility tier visualization for fish.

impact on the overall system. A key issue with this simulator is that it is possible to set the disaster threshold so high that the islands are wiped out quickly and the CPR threshold so low that the islands are essentially indestructible. So we need to find a setting that makes survival possible, but not certain, so that this system may operate in a "corridor of uncertainty." Therefore some parameters have been experimentally determined to provide a setting in which survival is possible but not guaranteed, unless the self-organizing mechanisms are used. This provides a "stable" context to investigate the relationships between independent and dependent variables.

5 Experimental Results

In this section, we describe the experimental results of four "survival trials." After we make a methodological observation on experimental design for such trials in section 5, each of the four subsequent subsections explores the performance of the multiagent system under varying degrees

of uncertainty, numbers of starting islands, and availability of self-organizing mechanisms. The aim is to investigate, in turn, the following four hypotheses:

Hypothesis A. A minimum number of starting agents, termed 'critical mass,' is required for the self-organizing mechanisms to have a positive effect (Trials A1–A6).

Hypothesis B. Increasing dimensions of uncertainty and complexity require increasing opportunity for self-organization (Trials B1–B4).

Hypothesis C. The various self-organizing mechanisms can interact to produce positive synergies (Trials C1–C4).

Hypothesis D. The various self-organizing mechanisms can interact to produce negative synergies (Trials D1–D2).

A Python3 program is used to generate the various plots of the survival probabilities in Figures 7–9. A run is regarded as successful if *all* agents survive for the entire duration of the simulation and the total quoted survival probability for each pair of agents and active organizations is averaged across $N = 30$ runs.

We look at the response to combinations of up to 12 islands with the eight linear combinations of trading (t) governance (g) and forecasting (f), representing their activity using the notation $t + g + f$ if all organizations are active. In this vein, a plot point of $t + f$ would represent active trading and forecasting, with governance deactivated.

Furthermore, the dimensions of uncertainty are denoted by: T, for a visible disaster period and CP for a visible common-pool threshold. Hence, the notation $CP + !\,T$ corresponds to a simulation with both the common-pool threshold visible and time period of disaster not visible.

We also define a set of parameters that remains constant across all simulations. Each simulation runs for 50 turns. We first model the disaster with a nonstochastic period of $T = 5$, with magnitude defined by a Gaussian distribution of mean of 6.5. The resultant damage felt scales this magnitude by 85. The common-pool threshold needed for the disaster to be mitigated is fixed at 300, with its initial resources set at 600. For an agent to "die," they must have *at most* 200 resources for five consecutive turns, starting with an initial 1,000 resources.

All actions carried out by the IIGO yield a cost of two resources, with the exception of broadcasting taxation. This is done to ensure that taxation can always be carried out despite an empty common pool, as otherwise, an empty pool can never be refilled. For this simulation, all rules are instantly put in play to increase the power of the governance organization.

Section 5.6 contains a summary of the entanglement of games and mechanisms and their interrelationships. It also indicates how qualitatively similar the complex scenario studied here is to those studied in cybernetics. Some of those principles and methods might be useful in further work, for example, the concepts of *stability* (Ashby, 1952) and of a *viable system* (Beer, 1972), because implicitly, one aim of the mechanisms for cooperative survival is to find a balance between control and effectiveness.

5.1 Methodology and Experimental Design

Owing to the large number of parameters involved in this simulator, it is unfeasible to test all parameter-value combinations using a sweep, for example, as all parameters are effectively unbounded. The methodology for deriving the parameters for the results reported in this section is based on the mean number of islands seen during experimentation. Parameters are selected to allow six agents to survive when at least two mechanisms are active and remain constant throughout the experimentation. Effectively, the parameters are fine-tuned for the case of relatively high survivability with a mean number of islands and subsequently used for exploring other settings.

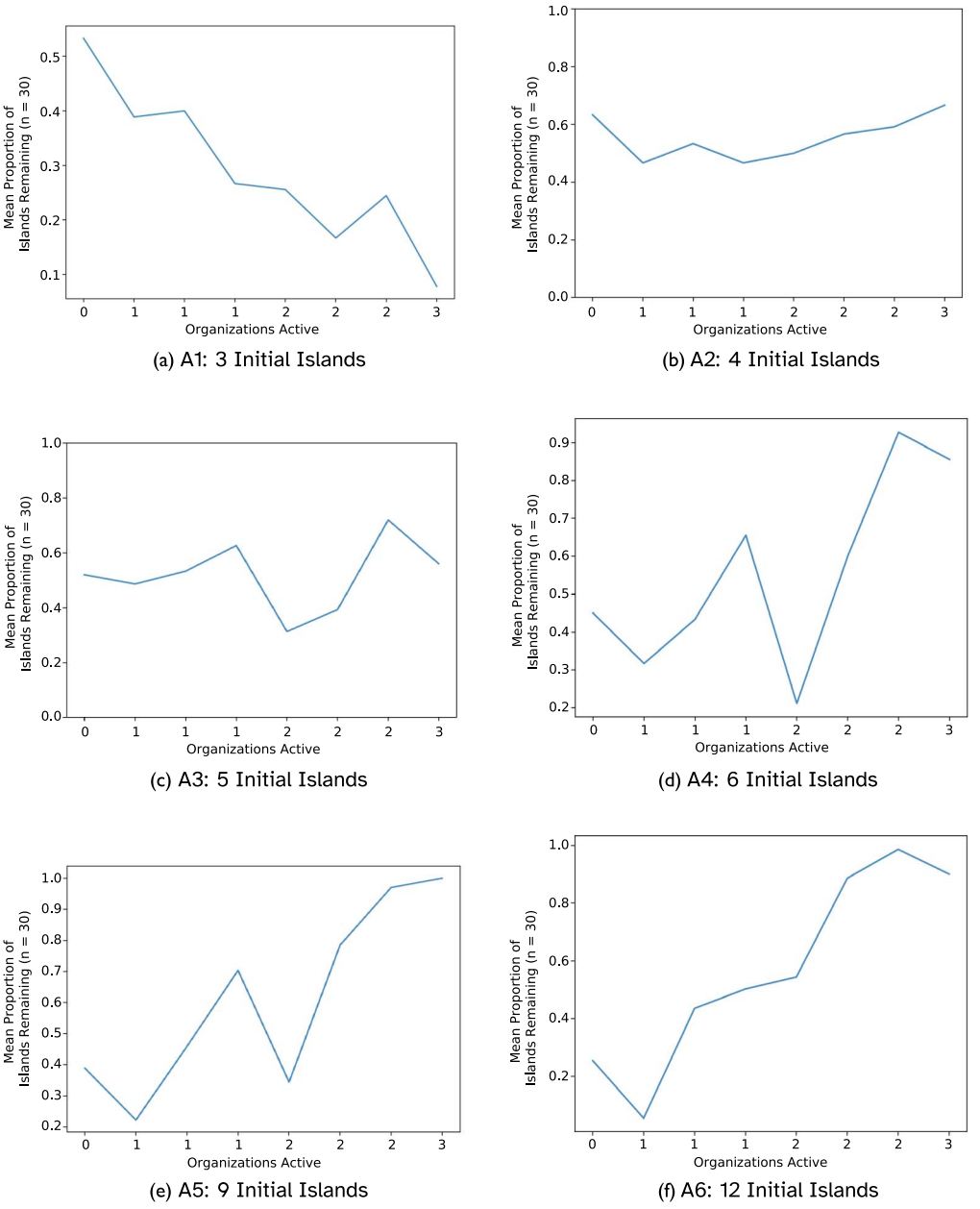


Figure 7. Survival probabilities for linear combinations of self-organizing mechanisms and starting islands for different degrees of uncertainty.

In general, when trying to analyze and understand a collective action situation, such as cooperative survival, one approach is to make the situation tractable by abstraction, assumption, and simplification. The problem with this approach is that while the situation can then be analyzed mathematically, it can create specious paradoxes. For example, Binmore (2005) likens the apparent paradox of noncooperation inherent to the common formulation of the prisoner's dilemma to throwing a strong swimmer into a lake with weights attached to their legs, then neglecting to mention the weights when claiming that it is paradoxical for strong swimmers to drown in this lake.

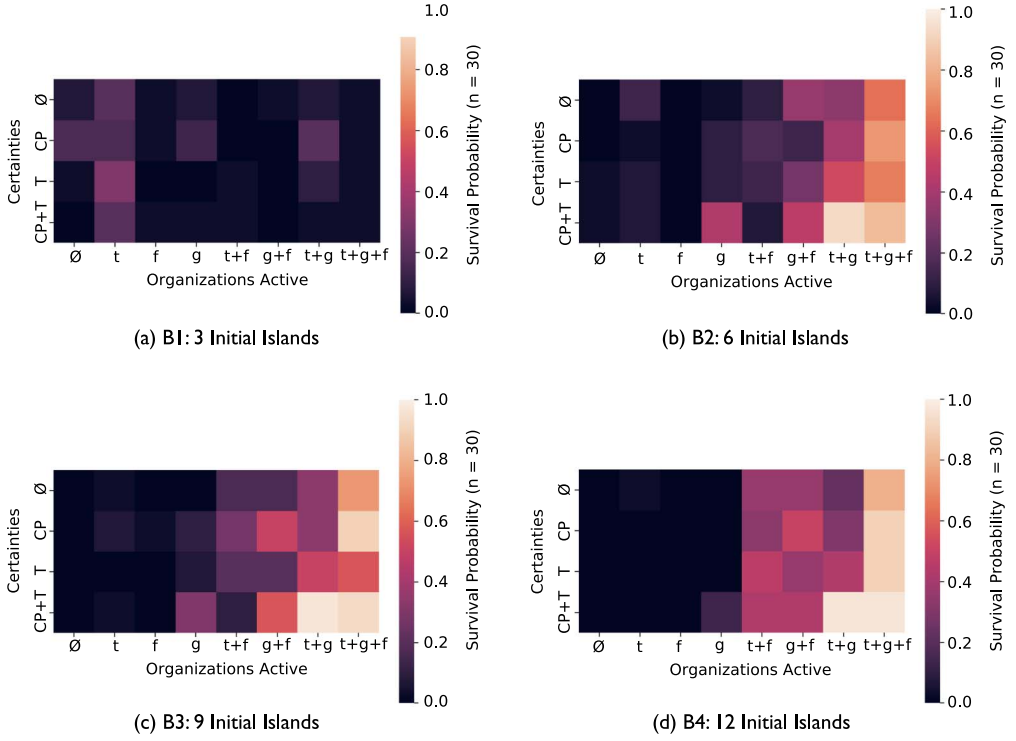


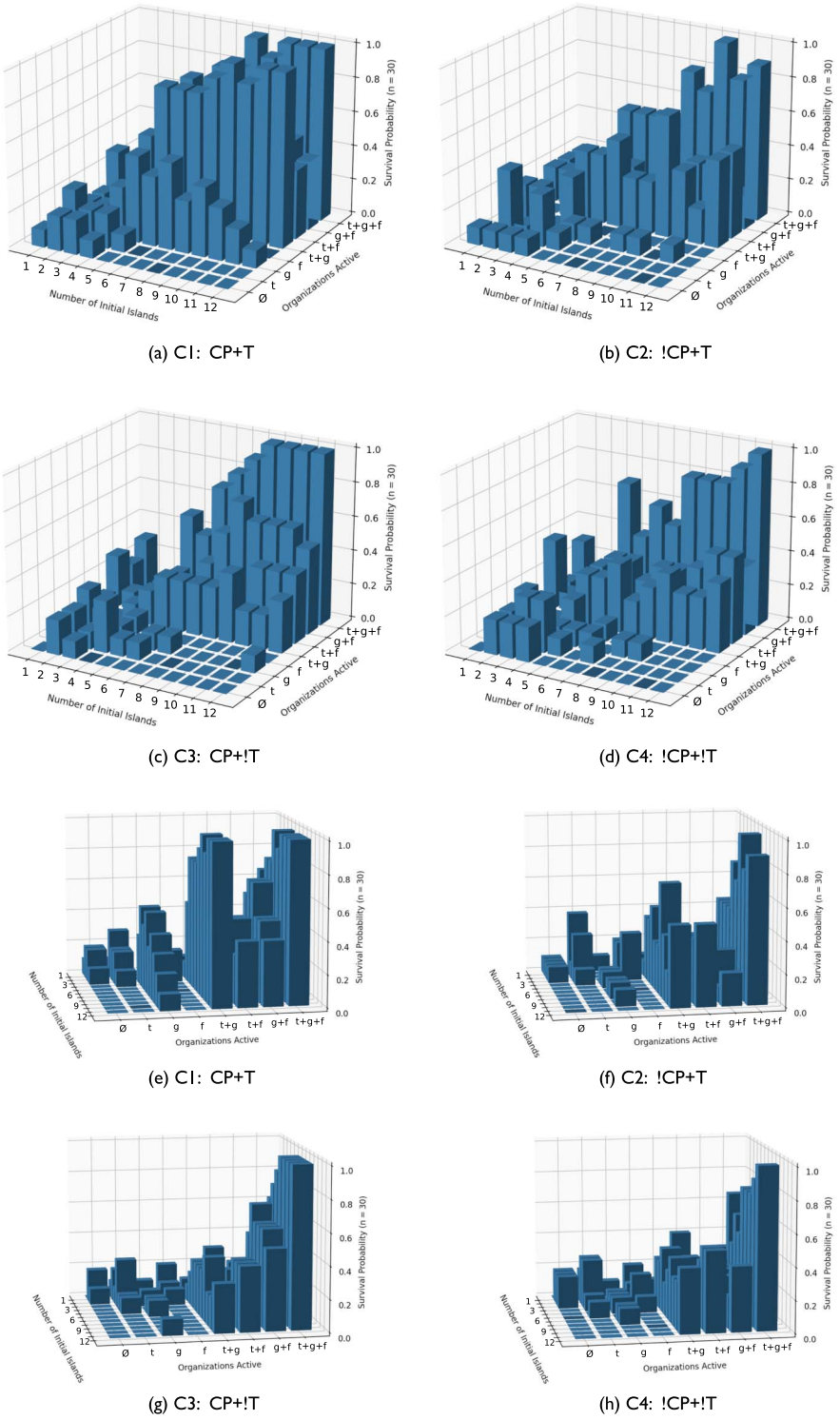
Figure 8. Survival probabilities for linear combinations of self-organizing mechanisms and degrees of uncertainty for varying numbers of starting islands. CP = visibility of the common pool threshold; T = visibility of the disaster period; t = trading; f = forecasting; g = governance.

An alternative approach is to recognize the interconnectedness and interdependence of features and factors in such situations that, together with randomness, nonlinearity, and so on, render the situation resistant to numerical analysis, and to use simulation instead (noting further that complementary approaches can offer different insights into different questions; Powers et al., 2018). However, trying to model the situation, and the agents in their entirety, means building a system with many parameters with many values, making a full exploration of a huge space by exhaustive parameter sweep computationally prohibitive, as well as making it much harder to identify and extract significant or meaningful relationships.

However, in *Just Six Numbers*, astronomer Martin Rees (2014) identifies just six physical constants that make for "meaningful" natural science and whose values, if they were only slightly different, would cause the universe to be completely uninteresting, in the sense that neither people nor planets, stars nor galaxies, would, or even could, exist. There certainly would not be any interesting behaviors, let alone people to propose and test theories about them.

Microworld simulations and experiments, of the kind investigated here, actually have similar properties. There are actually three classes of parameter: "constants," independent control variables, and dependent variables. Methodologically, the simulator's first job might be to identify and fine-tune a set of microworld constants, akin to the physical constants of the "real" universe. These provide sufficient underlying stability, which enables a systematic investigation of the relationship between the control and dependent variables that enable the simulator to validate, or otherwise, an experimental hypothesis.

Unfortunately, it can appear that this fine-tuning process involves assigning some perhaps seemingly arbitrary values to these physical constants, or, worse, cherry-picking results. Although we would not equate fine-tuning with cherry-picking, we should heed Ostrom's warning about basing the formulation of policy on numerical analysis or laboratory studies, which are far removed



(a) C1: CP+T (b) C2: !CP+T

(c) C3: CP+!T (d) C4: !CP+!T

(e) C1: CP+T (f) C2: !CP+T

(g) C3: CP+!T (h) C4: !CP+!T

Figure 9. Survival probabilities for linear combinations of self-organizing mechanisms and starting islands for different degrees of uncertainty.

from empirical experience, and we should also avoid ill- or unconsidered formulation of policy for high-stakes cooperative survival games based solely on these simulation results, under these conditions.

5.2 Experiments A: Need for Critical Mass

This set of experiments investigates the importance of a "critical mass" of survivors to facilitate group survival. We divide the experimentation into four sets, looking at the linear relations between the proportion of surviving islands and the availability of self-organizing mechanisms for 3, 6, 9, and 12 initial islands, while in an environment of *total certainty* (both common-pool threshold and disaster period are visible).

The results from this line of experimentation show that a threshold number of islands must be met to have any level of self-organizing mechanisms act favorably. Starting with Figure 7a, Trial A1 shows that any introduction of self-organizing mechanisms performs *worse* than having none at all. This illustrates that purely random behaviors (those that are uninfluenced by any form of "intelligent" self-organization) outperform the behaviors resulting from the available mechanisms.

This phenomenon can be explained through both the presence of "costly actions" and the demand for taxation. When the IIGO session is run, the various proceedings (such as appointing roles or proposing rules) have an associated cost that diminishes the common pool. Given the small mass of islands, very few cooperative metagames can be played, resulting in limited resource generation. With lessened global utility, the proportion of resources lost due to investment into the self-organizing mechanisms is far greater than it is with more starting islands.

In addition to this, the algorithm implemented by the IIGO *President* requires all islands to split the requirements for filling the common pool by both the number of islands and the number of days until disaster strikes. For this reason, the daily contribution needed for each island (assuming a desire to avoid sanctioning) grows *exponentially* as the number of alive islands decreases. For this reason, we conclude that, given an insufficient starting "mass," survival is improbable and unassisted by an introduction of survival mechanisms. We further note that the proportion of survivors is necessarily inflated, given that having even a single island survive represents a high proportion relative to the other trials.

Trials A2 and A3 (Figure 7b and 7c, respectively) show the behaviors of four and five starting islands. In both cases, there is a neutral impact from greater self-organization, resulting in no correlation between the proportion of survivors and the activity of the different mechanisms.

When starting with six islands, however, as in Trial A4 (Figure 7d), we see that there is a drastic positive impact from having more self-organizing mechanisms active, with this trend continuing through 9 and up to 12 islands in Trials A5 and A6 (Figure 7e and 7f, respectively). This confirms that there is a "critical mass" of islands that must be reached to facilitate a high proportion of survivors, along with a positive impact from increasing interference from self-organizing mechanisms. For this situation, we would assert that six islands represents the critical mass of survivors needed for a feasible system, although this result is not necessarily generalizable.

5.3 Experiments B: Increased Complexity

This second set of experiments aims to illustrate that an increase in uncertainty and complexity necessitates an increase in the number of opportunities for self-organization. To prove this, we vary two parameters in the simulator, the visibility of the common-pool threshold and the visibility of the disaster period, and produce plots showing variations in the number of starting islands, the availability of self-organizing mechanisms, and the probability of all islands surviving for each case.

An observation that can be made from all four heat maps, irrespective of the level of certainty, is that for an increasing number of islands, it is near impossible to survive without any organizations active. We propose three possible reasons why a low number of islands may be able to survive without the ability to self-organize (albeit in very few cases), as shown by Trials B1 and B2 (Figure 8a and 8b, respectively).

First, owing to a relatively high initial common pool, because the initial resources in the common pool stay fixed across all survival trials, a smaller multiagent system means that each agent can take a larger share of the common pool without completely draining it.

Second, because the resource reduction of the common pool is calculated based on the total damage felt across the multiagent system (i.e., the sum of the damages of each individual agent), having fewer islands results in a lower overall reduction of resources and helps maintain the size of the common pool.

Finally, the geographical distribution of islands sees them as uniformly distributed points around a ring. Because we have fewer islands, the average distance between islands increases, meaning that the average distance from each island to the disaster will increase. This again results in less damage being felt and hence in a lower impact on the common pool. As the number of islands increases, we enter an economy of scarcity: Despite an increase in returns from foraging due to the increased number of foragers, there are simply not enough self-organizing mechanisms in play to help redistribute the resources. Having run further diagnostics, however, the variation from 3 to 12 islands yields an average distance decrease of approximately 20%, which we assert is insignificant in comparison to the change in the number of islands.

Considering in detail the role of interdependent mechanisms (over mechanisms in isolation) with increasing uncertainty, we look at the disparity between the survivability given three active organizations over the survivability given two or fewer. Having reached the critical mass of islands established in Experiments A, we have up to a 100% survival chance with three and between 60% and 80% with just two. Trial B4 shows three transitions in color variation, where a surge from a 10% to 60% to 100% survival rate is achieved by introducing subsequent survival mechanisms.

Supporting this conclusion is the change in color between the number of mechanisms with respect to the change in uncertainty. Looking at Trial B3 (Figure 8c), we see far greater leaps in lightness with decreased certainty. Quantitatively, with a certainty of "none," we see that a survival rate of ≈20% rapidly increases to ≈80%, yielding a 60% increase from two to three mechanisms. We contrast this with the full-certainty case, in which the increase is from ≈70% to 90%, yielding a mere 20% increase. This pattern can be seen across graphs for Trials B2–B4 (neglecting the case below the critical mass) in Figure 8b–8d, from which we conclude that, with increasing uncertainty, there is a greater reliance on higher levels of interdependence.

5.4 Experiments C: Positive Synergies

This penultimate set of experiments aims to illustrate how, despite the general trend that an increased number of self-organizing methods tends to yield an increase in survivability, the choice of mechanisms must be carefully considered, and such mechanisms may have to interact in mutually self-reinforcing ways. This set of experiments is visualized with linear combinations of self-organizing mechanisms and starting islands for different degrees of uncertainty, achieved by varying the visibility of the common-pool threshold and disaster period. For the plots in Figure 9, we denote the visibility of the common-pool threshold and disaster period as *CP* and *T*, respectively.

The aim of this experiment was to investigate the importance of forecasting and to hypothesize that when all disaster parameters are known, this organization offers no additional support for survival. The first contour plot for Trial C1, Figure 9a, shows the instance of complete certainty when both common-pool threshold and disaster period are known. In this plot, we see that survival probability is optimized with either all organizations active or simply just the trading and governance organizations active. The valley caused by the inactivity of governance shows that rules for taxation and resource management *must* be in place for the common pool to be regenerated and appropriately mitigate disaster. Given the situation of having perfect knowledge of disaster and how to mitigate it (through the common-pool threshold), forecasting offers no benefits to survivability.

The general trend from Trials C1–C4 (in Figure 9, seen from the front [Figure 9a–9d] and from the side [Figure 9e–9h]) is that the disparity between the survival probabilities of three mechanisms and two mechanisms *increases*. It can therefore be inferred that increased uncertainty requires

increased levels of self-organization, as the same level of survivability cannot be achieved between trials without introducing additional means of self-organization. Importantly, a perfect 100% survival rate is feasible across all degrees of uncertainty, so long as there is a sufficiently high critical mass and all three mechanisms are active.

The broader trend from plots for Trials C1–C4 (Figure 9a–9d) indicates that the presence of two self-organizing mechanisms yields widely varying results, depending on which two mechanisms are active at any one time. Drawing on Trial C3 specifically, we can see a disparity between the performances of $t + f$ and $g + f$, where $g + f$ drastically outperforms the former. This relation demonstrates that although forecasting is an important mechanism for informing taxation, this knowledge is useless (c.f. Dillon, 2010) in the absence of an institutional power to enforce said taxation. There is no point being an immaculate prophet of doom if no one believes a word or is willing to do anything about it.

A similar relation can be seen in Trial C2 regarding $t + g$ and $g + f$. Although there may be sufficient institutional power to enforce taxation, the stochastic nature of the stag hunt means that taxation quickly becomes unfair when resources cannot be redistributed, and a form of oligarchy emerges.

Ultimately, the power of each self-organizing mechanism is amplified when accompanied by other, mutually reinforcing mechanisms. This is particularly prevalent when only a subset of the full range of self-organizing mechanisms is available, as results varied drastically based on the pairings of mechanisms in the absence of the third. In the following Experiments D, we further investigate the importance of mechanism selection, as other negative synergies may emerge, such as the superfluous signaling identified here.

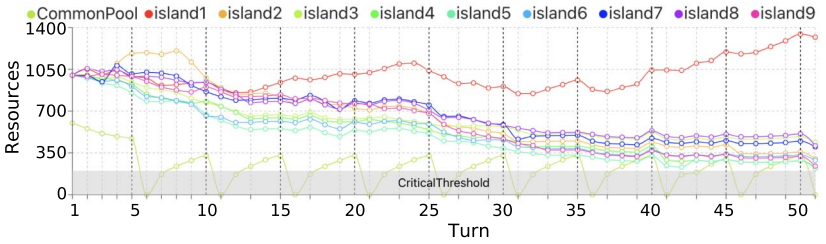
5.5 Experiments D: Negative Synergies

Although the overarching trend in all previous experiments, Trials A–C, affirms that an increasing number of self-organizing mechanisms is conducive to survival, an important criticism to identify is that these plots are an *aggregation* of multiple results and offer no insight into what happens "under the hood"; they merely show the broader picture. For this reason, we use this section to illustrate instances in which self-organizing mechanisms may function in *pernicious* ways by drawing specific attention to the amount of time an island spends in positions of power and the overall resources it possesses throughout the simulation.

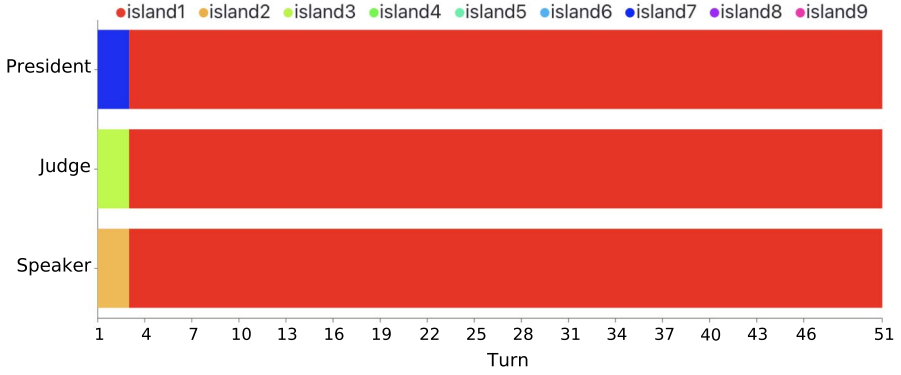
Trial D1 shows a clear case of the *iron law of oligarchy*, where a purely democratic system has the capability to degenerate into an oligarchical system. This is suitably evidenced through the poor resource distribution in D1, in which all islands other than Island 1 are forced to live equally, but the lion's share of resources is withheld (Figure 10a). The explanation for this phenomenon can be seen in Figure 10b, as Island 1 is able to occupy all roles in parliament. This means that the island is immune to any form of sanction and receives all additional salaries from the IIGO.

Island 1 is able to remain in power because of the absence of the trading mechanism, meaning that a well-established social network is unable to be formed. Owing to the absence of trade, a case of individuality is further created by means of positive feedback, as wealth cannot be redistributed. This experimentation hence demonstrates that, for a fair system, governance must be activated in parallel with trade, not only to facilitate strong links between the nodes in the social network but to redistribute an imbalance of wealth.

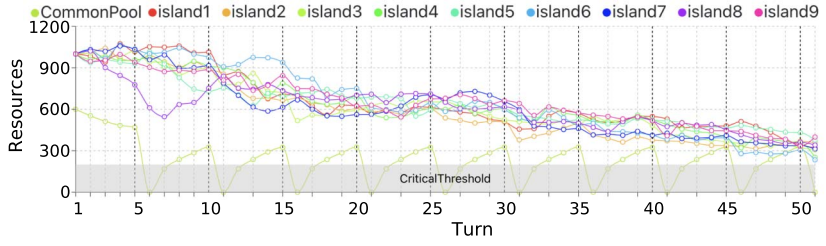
We contrast this performance to that of the simulation in Trial D2, where trade is activated to allow for intelligent social network formation. Figure 10c shows a "fairer" distribution of resources, in which all islands have a similar, if not equal, quantity of resources at each turn of the simulation. We also identify a less unequal spread of resources when contrasting Figure 10a and 10c, and a wider spread of institutionalized powers (comparing Figure 10b and 10d), showing that trade is heavily influential in equalizing resources across a social network. Furthermore, the distribution of positions of power in the IIGO is far more randomized, again demonstrating the power of opinion formation through trade in producing a "fair" system.



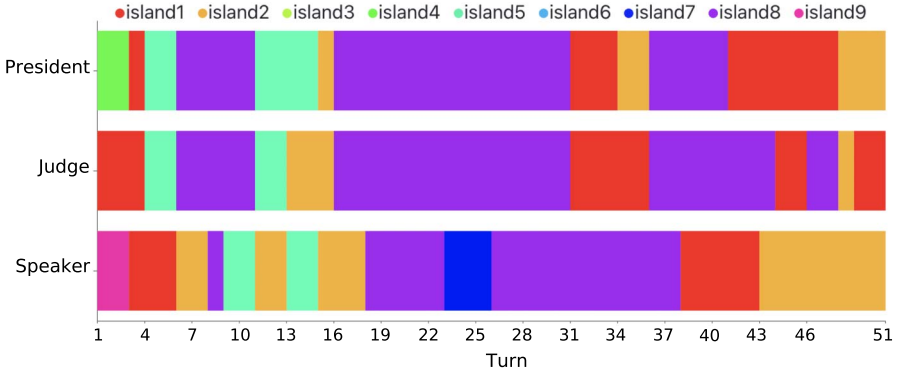
(a) D1: Resources Held



(b) D1: Roles in IIGO



(c) D2: Resources Held



(d) D2: Roles in IIGO

Figure 10. Distribution of resources and time spent in positions of institutionalized power (roles) in exemplar trials.

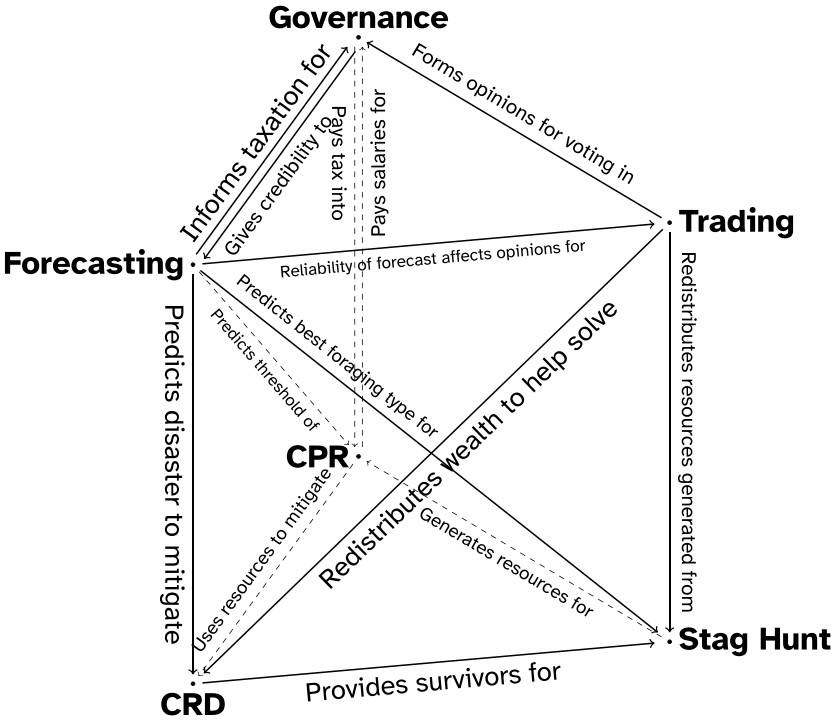


Figure 11. Visualization of interdependencies between organizations and games. CPR = common-pool resource; CRD = collective risk dilemma.

Importantly, both Trials D1 and D2 are successful survival trials, with the average distribution of resources fairly similar in both cases, barring the dominance of Island 1 in Trial D1. This shows that, while the concept of an oligarchy stands in direct opposition to a "fair democracy," it is not necessarily disadvantageous for the collective. For this reason, we term the emergence of a single powerful island as an "enlightened dictatorship."

We also see pernicious interactions between the self-organizing mechanisms by referring back to Trial A1, where the lack of islands in the archipelago results in overtaxation, ultimately limiting the prospect of survival. This demonstrates that the choice of activation of self-organizing mechanisms must be carefully made, relative to both their pairings and the number of islands. This is further supported by inspection of Figure 11, in which the interactions between games and self-organizing mechanisms is visualized. Removing any one of these nodes massively affects the cyclical nature of the interdependencies, which can lead to unexpected effects.

Overall, we consider this set of experiments to be a culmination of the previous three sets, as it unites the key principles from each. Increased uncertainty necessitates increased opportunities for self-organization, provided this set of self-organization mechanisms is carefully constructed and underpinned by a critical mass of agents. Otherwise, it is entirely possible that these mechanisms will have no positive influence, potentially leading to instances of oligarchy and superfluous signaling.

5.6 Summary of Experiments: Entanglement

The overall conclusion that we draw from these experimental investigations is that there are complex dependencies between the subgames—and the self-organizing mechanisms are intended to provide political metagames to help solve corresponding subgames—but that there are also complex dependencies between the political games. It would seem that there is a high level of interdependence between the subgames and self-organizing mechanisms but also intradependence between the subgames and intradependence between the self-organizing mechanisms, as illustrated in Figure 11.

This entanglement of subgames and self-organizing mechanisms between themselves and each other reinforces the importance of reflection (Landauer & Bellman, 2016), self-awareness (Lewis et al., 2016), and introspection (Holland et al., 2013); their role in reflective governance both in socioecological systems (Dryzek & Pickering, 2016) and in socio-techno-ecological systems (Pitt et al., 2020); and the importance of cybernetic concepts of stability and viability in self-organizing systems (Ashby, 1952; Beer, 1972). A key point of systemic reflection is for components of the system themselves to balance the tensions between systemic drivers pushing the system to optimize for polar-opposite values. A similar requirement is observable here: to maintain the tension between three points (at the level of subgames and the level of self-organizing mechanisms) and between points at different levels.

6 Related and Further Work

6.1 Related Work

We envision this article as helping to unify various other pieces of work concerning variations in CPR management: LPG games. In this section, we explore thematic connections to this research concerning timing uncertainty and its effect on collective risk (Domingos et al., 2020), institutional rules for reward and sanctioning (Powers, 2018), and the use of a common pool for disaster prevention (Milinski et al., 2008).

The general theme with these works is that they investigate in greater detail some of the components that make up this simulation software. We see this work as a unification of simplified versions of the related literature, made with the intention of offering insight into an *interdependent* system. For this reason, while elements of the related work stand present in this article, they serve instead to build up a more complex, interrelated system.

Looking first at Domingos and colleagues' (2020) work on timing uncertainty, we draw on the main parallel that a fair criticism of conventional LPGs and CRDs is the limited relevance to real-world scenarios given the a priori knowledge of the period of disaster and needed threshold. Domingos and colleagues hence offer a simulation where timing is uncertain and observes the results of an otherwise conventional LPG. The way in which this article differs is twofold; as well as introducing a second component of uncertainty (through an unknown threshold), we take a different approach to what role the purpose of timing uncertainty plays. Domingos and colleagues' work investigates how decision-making is affected by timing uncertainty, by creating various strategies for common-pool contribution. We instead utilize timing uncertainty purely as a means of increasing the overall uncertainty of the system to evaluate the performance of the different self-organizing mechanisms, giving one fixed contribution method based on the time remaining and number of agents alive. For this reason, we conclude that strategic contribution is needed through self-organization, as randomized taxation is an insufficiently powerful technique.

This article also mirrors Powers's (2018) work, which is predicated on Ostrom institution theory and how institutional rules can promote cooperation in economic activities. Powers's paper references the fact that both the sharing of information (of other agents) and coordinated systems of reward and punishment facilitate cooperation in a social network and can help influence how economic interactions occur. From this, Powers provides a model of sanctioning institutions to help define explicitly the evolution of institutional rules, looking at the convention of rewarding cooperators and sanctioning (punishing) defectors. We take Ostrom's and Powers's insights into this topic to utilize a system of rules and sanctioning, however, with great simplification. Rule selection in the case of this article is trivial, putting all predefined rules in play. Although the simulator allows for rule *selection*, we again propose a simplistic strategy of random selection with guaranteed approval. This reflects a trivial institutional rule implementation process with little to no evolution. For this reason, we assert that the rule set in this simulator serves only to constitute the governance organization, effectively creating a simulator with switchable rule activation. Hence we conclude that sanctioning, taxation strategy, and rules are required for a feasible level of survivability and tie these attributes together under the umbrella term of *governance*.

Finally, through investigating Milinski and colleagues' (2018) work, we see a different approach to solving a CRD of CPR management. This work looks at the strategy of convincing agents that a failure to invest enough into the common pool is very likely to cause grave financial loss to the individual. Parallels between this work and Milinski and colleagues' are visible, particularly in the way that the resource deduction from each agent's private pool is controlled. Assuming that the threshold is not met, we use an output scalar to greater deplete the agents' resources, resulting in a similar affirmation that a failure to mitigate disaster results in a more severe personal loss. Again, these works bifurcate because of the nature of the problem we are looking to solve—we propose that the collective risk dilemma *in contention with two other metagames* can be solved through a series of interdependent self-organizing mechanisms, not just a fixed strategy.

6.2 Further Work

Although we have supplied three feasible self-organizing mechanisms for approaching the problem of intertwined subgames, it becomes an open question as to whether other socially constructed mechanisms could also work effectively. This creates an opportunity for further investigation. Additionally, experiments could be conducted over more iterations for reproducibility or with a "system of systems," where we envision a "super-archipelago" comprising many other archipelagos.

The complexity of this simulation software gives rise to multiple possibilities for future work. The first point to investigate is the prospect of randomizing the disaster frequency, instead giving a fixed disaster period. This will greatly reduce the power of the current forecasting algorithm (as the first guess that is made is the correct guess, because it is equal to the first turn that disaster strikes), leading to increased randomness in the simulator. In addition, the algorithm for this case could be implemented similarly to the threshold estimation, where a running prediction can be increased and decreased according to the turn on which disaster strikes.

As well as this, the forecasting organization could be improved to aim to replicate a form of "foreign aid": by regressing to the average location of disaster, forecasting could be used to help predict which agent is likely to be the worst affected during the next disaster. This knowledge could subsequently be used during trading to give a burst of support to the agent in danger for further mitigation.

A means of improving the reliability of the forecasting organization (to avoid the fate of Cassandra alluded to in Experiments D) is through forming a social network influenced by the quality of forecasting. Through repeated successful predictions, it can be inductively reasoned that this will give rise to "experts" with high social standing that are acclaimed for reliable forecasting. A weighting can subsequently be applied proportionally to the level of expertise to aggregate a final, accurate prediction.

Surplus to this, the current formulation of the simulator naturally gives rise to a machine learning approach in many aspects. Regression algorithms can be applied to the role of *President* to determine the optimal level of taxation, or a classification algorithm can be used to define the various rules as beneficial or not. Naturally, any aspect of the simulator that introduces the possibility of choice may be solvable through modern machine learning methods; however, it may limit the system's heterogeneity.

7 Summary and Conclusions

In summary, the contributions of this article are threefold:

1. We have defined an innovative analytical and experimental scenario for exploring cooperative survival games with four parameters: scale, uncertainty, complexity, and opportunity for self-organization.
2. We have designed, specified, and implemented a self-organizing multiagent system with new algorithms for three self-organizing mechanisms: trading, forecasting, and governance.

3. We have conducted a series of experiments that have shown the following:

   - Complex cooperative survival games are solvable given a critical mass of survivors and some opportunity for self-organization.
   - Even with critical mass, increasing dimensions of uncertainty and complexity require more and "better" opportunities for self-organization.
   - However, self-organizing mechanisms can interact in mutually supporting ways but also in unexpected ways with potentially pernicious outcomes.

In more detail, the general trend observed in all the survival trials is that, irrespective of uncertainty, the survival chance increased as more self-organizing mechanisms were made available. The main exception was in the case of total certainty, in which the presence of forecasting offered no greater chance of survival than without: It would seem, like Cassandra, that there is no point being an immaculate prophet of doom if no one believes a word or is willing to do anything about it, in the absence of either some form of coercive authority (Olson, 1965) or some decentralized incentive to collective action (Ostrom, 1990).

This observation indicates that while the three subgames of the cooperative survival game are interdependent, the self-organizing mechanism designed to help solve each subgame does not work in isolation either, and these mechanisms can be mutually supporting. The experiments showed that introduction of more self-organizing mechanisms can provide increasing benefit, as the interplay between such mechanisms allows for a greater range of possible actions and hence multiple ways of assessing the different metrics in the social system. A particular example of this is how the friendship value shared between agents may be initially formulated through trade; however, this was reevaluated through their willingness to broadcast taxation in government.

Furthermore, increased uncertainty yielded an overall lower survival chance, irrespective of the number of active mechanisms; however, there were most instances when a 100% survival chance was maintained when all three mechanisms were active. This showed that the self-organizing mechanisms are sufficient for allowing cooperative survival throughout all levels of disaster uncertainty; comparatively, though, the performance of all three mechanisms, when active, relative to one acting alone, is much greater than the sum of its parts.

However, while an increase in uncertainty necessitates an increase in opportunities for self-organization for a successful solution, the experiments also demonstrated that interplay can be potentially pernicious and furthermore required the active participation of sufficient agents for the self-organizing mechanisms to ensure a positive rather than a pernicious influence (i.e., even the subgames required a critical mass of "players," as well as the overall cooperative survival game requiring a critical mass of survivors).

In conclusion, then, the introduction of multiple self-organizing mechanisms to complex and *sensitive* situations should be carefully designed to ensure that such mechanisms are mutually supportive—with the strong caveat that the intended beneficial and prosocial relationship is not always instantly obvious and should not be taken for granted. However, we would contend that this is precisely the kind of complementary insight that ALife simulations of the kind described here can bring to applications like statistical epidemiology for public health policy making, population modeling for urban planning, and legal modeling for informing legislative drafting. We need to know what happens when people are empowered with tools to reshape their own environment, and what happens (such being the power of generativity) when they reshape the tools in ways that the designers never intended or even imagined.

Because it is always possible to game the metagame, effective processes for reflection in collective self-governance will be critically important in the future. Reflection is, loosely speaking, the idea that in control systems, some components have an internal model of the system (including themselves), can reason about its (the system's and the component's) behavior and performance, and can adjust or reconfigure some of the control variables or parameters accordingly (Ashby, 1952;

Landauer & Bellman, 2003). The need for such reflection has been identified for self-governance, has been studied computationally in diverse contexts such as self-improving systems for systems integration (Bellman et al., 2014), and has motivated the need for run-time models (Landauer & Bellman, 2016). The idea of reflection has also featured in recommendations for the collective self-governance of socioecological systems (Dryzek & Pickering, 2016) and algorithmic self-governance of socio-techno-ecological systems (Pitt et al., 2020). Especially in those latter contexts, where sustainability is such a key feature, the current study highlights the need for reflection not just as a matter of ALife but as a matter of ALife and ADeath as well.

Acknowledgments

We are especially grateful to the anonymous reviewers for their many helpful and insightful comments. This work is partly based on the simulator developed during the COVID-19 pandemic by the Imperial College Department of Electrical and Electronic Engineering SOMAS cohort, 2020–2021.

References

Artikis, A., Sergot, M. J., & Paliouras, G. (2015). An event calculus for event recognition. *IEEE Transactions on Knowledge and Data Engineering*, *27*(4), 895–908. https://doi.org/10.1109/TKDE.2014.2356476

Artikis, A., Sergot, M., & Pitt, J. (2009). Specifying norm-governed computational societies. *ACM Transactions on Computational Logic*, *10*(1), 1–42. https://doi.org/10.1145/1459010.1459011

Ashby, W. (1952). *Design for a brain*. Chapman Hall.

Beer, S. (1972). *Brain of the firm*. Allen Lane.

Bekius, F., Meijer, S., & Thomassen, H. (2022). A real case application of game theoretical concepts in a complex decision-making process: Case study ERTMS. *Group Decision and Negotiation*, *31*(1), 153–185. https://doi.org/10.1007/s10726-021-09762-x

Bellman, K., Tomforde, S., & Würtz, R. (2014). Interwoven systems: Self-improving systems integration. In *Eighth IEEE International Conference SASO Workshops* (pp. 123–127). IEEE. https://doi.org/10.1109/SASOW.2014.21

Berger, P., & Luckmann, T. (1966). *The social construction of reality: A treatise in the sociology of knowledge*. First Anchor Books.

Binmore, K. (2005). *Natural justice*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195178111.001.0001

Bourazeri, A., & Pitt, J. (2018). Collective attention and active consumer participation in community energy systems. *International Journal of Human-Computer Studies*, *119*, 1–11. https://doi.org/10.1016/j.ijhcs.2018.06.001

Bowles, S., & Gintis, H. (2011). *A cooperative species*. Princeton University Press. https://doi.org/10.23943/princeton/9780691151250.001.0001

Brennan, G., & Pettit, P. (2005). *The economy of esteem*. Oxford University Press. https://doi.org/10.1093/0199246483.001.0001

Briggs, J. (1970). *Never in anger: Portrait of an Eskimo family*. Harvard University Press.

Broad, W. J. (2007). *The oracle: Ancient Delphi and the science behind its lost secrets*. Penguin.

Byrne, J. P. (2006). *Daily life during the Black Death*. Greenwood.

Cadsby, C. B., & Maynes, E. (1999). Voluntary provision of threshold public goods with continuous contributions: Experimental evidence. *Journal of Public Economics*, *71*(1), 53–73. https://doi.org/10.1016/S0047-2727(98)00049-8

Carlsson, H., & van Damme, E. (1993). Equilibrium selection in stag hunt games. In K. Binmore, A. Kirman, & P. Tani (Eds.), *Frontiers of game theory* (pp. 237–254). MIT Press.

Dillon, M. (2010). Kassandra: Mantic, maenadic or manic? Gender and the nature of prophetic experience in ancient Greece. In *Proceedings of the AASR Conference, University Of Auckland, New Zealand July 6–11, 2008*. Australian Association for the Study of Religions. https://openjournals.library.sydney.edu.au/AASR/article/view/2341

Domingos, E. F., Grujić, J., Burguillo, J. C., Kirchsteiger, G., Santos, F. C., & Lenaerts, T. (2020). Timing uncertainty in collective risk dilemmas encourages group reciprocation and polarization. *iScience*, *23*(12), Article 101752. https://doi.org/10.1016/j.isci.2020.101752, PubMed: 33294777

Dryzek, J., & Pickering, J. (2016). Deliberation as a catalyst for reflexive environmental governance. *Ecological Economics*, *131*, 353–360. https://doi.org/10.1016/j.ecolecon.2016.09.011

Durant, W. (2006). *The story of philosophy*. Pocket Books.

Durmus, Y., Ozgovde, A., & Ersoy, C. (2011). Distributed and online fair resource management in video surveillance sensor networks. *IEEE Transactions on Mobile Computing*, *11*(5), 835–848. https://doi.org/10.1109/TMC.2011.115

Frey, S., & Sumner, R. W. (2018). *Emergence of complex institutions in a large population of self-governing communities*. arXiv:1804.10312. https://doi.org/10.48550/arXiv.1804.10312

Harrison, G. W., Martínez-Correa, J., Swarthout, J. T., & Ulm, E. R. (2017). Scoring rules for subjective probability distributions. *Journal of Economic Behavior and Organization*, *134*, 430–448. https://doi.org/10.1016/j.jebo.2016.12.001

Holland, S., Pitt, J., Sanderson, D., & Busquets, D. (2013). Reasoning and reflection in the Game of Nomic: Self-organising self-aware agents with mutable rule-sets. In *7th IEEE International Conference SASO Workshops* (pp. 101–106). IEEE. https://doi.org/10.1109/SASOW.2013.27

Jones, A., & Sergot, M. (1993). On the characterisation of law and computer systems: The normative systems perspective. In J.-J. Meyer & R. Wieringa (Eds.), *Deontic logic in computer science: Normative system specification* (pp. 275–307). Wiley.

Koomen, R., & Herrmann, E. (2018). An investigation of children's strategies for overcoming the tragedy of the commons. *Natural Human Behaviour*, *2*(5), 348–355. https://doi.org/10.1038/s41562-018-0327-2, PubMed: 30962602

Landauer, C., & Bellman, K. L. (2003). Meta-analysis and reflection as system development strategies. In D. L. Hicks (Ed.), *International Symposium on Metainformatics MIS* (pp. 178–196). Springer. https://doi.org/10.1007/978-3-540-24647-3_16

Landauer, C., & Bellman, K. L. (2016). Reflective systems need models at run time. In *Proceedings of the 11th International Workshop on Models@run.time* (pp. 52–59). ceur-ws.org. https://ceur-ws.org/Vol-1742/

Lansing, J. S., & Kremer, J. N. (1993). Emergent properties of Balinese water temple network: Coadaptation on a rugged fitness landscape. *American Anthropologist*, *95*(1), 97–114. https://doi.org/10.1525/aa.1993.95.1.02a00050

Lewis, P., Platzner, M., Rinner, B., Tørresen, J., & Yao, X. (Eds.) (2016). *Self-aware computing systems: An engineering approach*. Springer. https://doi.org/10.1007/978-3-319-39675-0

Malinowski, B. (1920). Kula: The circulating exchange of valuables in the archipelagoes of eastern New Guinea. *Man*, *20*, 97–105. https://doi.org/10.2307/2840430

Manville, B., & Ober, J. (2019). In search of democracy 4.0: Is democracy as we know it destined to die? *IEEE Technology and Society Magazine*, *38*(1), 32–42. https://doi.org/10.1109/MTS.2019.2894458

McAuliffe, K., Blake, P. R., Steinbeis, N., & Warneken, F. (2017). The developmental foundations of human fairness. *Nature Human Behaviour*, *1*(2), Article 0042. https://doi.org/10.1038/s41562-016-0042

Milinski, M., Sommerfeld, R. D., Krambeck, H. J., Reed, F. A., & Marotzke, J. (2008). The collective-risk social dilemma and the prevention of simulated dangerous climate change. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(7), 2291–2294. https://doi.org/10.1073/pnas.0709546105, PubMed: 18287081

Nikoloska, I., & Simeone, O. (2021). Fast power control adaptation via meta-learning for random edge graph neural networks. In *2021 IEEE 22nd International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)* (pp. 146–150). IEEE. https://doi.org/10.1109/SPAWC51858.2021.9593131

Norberg-Hodge, H. (1991). *Ancient futures: Learning from Ladakh*. Sierra Club Books.

Nowak, A., Vallacher, R., Rychwalska, A., Roszczyńska-Kurasińska, M., Ziembowicz, K., Biesaga, M., & Kacprzyk-Murawska, M. (2020). *Target in control: Social influence as distributed information processing*. Springer. https://doi.org/10.1007/978-3-030-30622-9

Ober, J. (2017). *Demopolis: Democracy before liberalism in theory and practice*. Cambridge University Press. https://doi.org/10.1017/9781108226790

Olson, M. (1965). *The logic of collective action*. Harvard University Press.

Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. Cambridge University Press. https://doi.org/10.1017/CBO9780511807763

Ostrom, E., & Ahn, T. K. (2003). *Foundations of social capital*. Edward Elgar.

Petruzzi, P. E., Pitt, J., & Busquets, D. (2017). Electronic social capital for self-organising multi-agent systems. *ACM Transactions on Autonomous and Adaptive Systems*, *12*(3), Article 13. https://doi.org/10.1145/3124642

Pitt, J. (2021). *Self-organising multi-agent systems: Algorithmic foundations of cyber-anarcho-socialism*. World Scientific. https://doi.org/10.1142/q0307

Pitt, J., Busquets, D., Bourazeri, A., & Petruzzi, P. (2014). Collective intelligence and algorithmic governance of socio-technical systems. In D. Miorandi, V. Maltese, M. Rovatsos, A. Nijholt, & J. Stewart (Eds.), *Social collective intelligence: Combining the powers of humans and machines to build a smarter society* (pp. 31–50). Springer. https://doi.org/10.1007/978-3-319-08681-1_2

Pitt, J., Dryzek, J., & Ober, J. (2020). Algorithmic reflexive governance for socio-techno-ecological systems. *IEEE Technology and Society Magazine*, *39*(2), 52–59. https://doi.org/10.1109/MTS.2020.2991500

Pitt, J., & Hart, E. (2017). For flux sake: The confluence of socially- and biologically-inspired computing for engineering change in open systems. In *2017 IEEE 2nd International Workshops on Foundations and Applications of Self* Systems (FAS*W)* (pp. 45–50). IEEE. https://doi.org/10.1109/FAS-W.2017.119

Pitt, J., Kamara, L., Sergot, M., & Artikis, A. (2006). Voting in multi-agent systems. *Computer Journal*, *49*(2), 156–170. https://doi.org/10.1093/comjnl/bxh164

Pitt, J., & Ober, J. (2018). Democracy by design: Basic democracy and the self-organisation of collective governance. In *12th IEEE International Conference SASO* (pp. 20–29). IEEE. https://doi.org/10.1109/SASO.2018.00013

Powers, S. (2018). The institutional approach for modeling the evolution of human societies. *Artificial Life*, *24*(1), 10–28. https://doi.org/10.1162/ARTL_a_00251, PubMed: 29369715

Powers, S., Ekart, A., & Lewis, P. (2018). Modelling enduring institutions: The complementarity of evolutionary and agent-based approaches. *Cognitive Systems Research*, *52*, 57–81. https://doi.org/10.1016/j.cogsys.2018.04.012

Putnam, R. D. (2000). *Bowling alone: The collapse and revival of American community*. Simon & Schuster. https://doi.org/10.1145/358916.361990

Rees, M. (2014). *Just six numbers*. Orion.

Robert, S., Robert, H., Evans, W., Honemann, D., & Balch, T. (2000). *Robert's rules of order* (Newly rev., 10th ed.). Perseus.

Surowiecki, J. (2004). *The wisdom of crowds*. Little, Brown.

Ti, B., Li, G., Zhou, M., & Wang, J. (2022). Resilience assessment and improvement for cyber-physical power systems under typhoon disasters. *IEEE Transactions on Smart Grid*, *13*(1), 783–794. https://doi.org/10.1109/TSG.2021.3114512