Incentivising Participation With Exclusionary Sanctions (Full)

Buster Blackledge[1], Antonios Papaoikonomou[1], Matthew Scott[1], Asimina Mertzani[1], Noan Le Renard[1], Hazem Masoud[1], and Jeremy Pitt[1]

Imperial College London, London SW7 2BX, United Kingdom
(bb922,ap3422,mss2518,s20,nml18, hm1822,jpitt)@ic.ac.uk

Abstract. Some cooperative survival situations require all members of a group to participate equally in collective action; however, if the only sanction for non-participatory free-riding is exclusion, then it can be ineffective, as exclusion is indistinguishable from non-participation. The question then is: how does a group, that can define a set of mutually agreed conventional rules, incentivise participation that supports collective survival when the only sanctioning instrument is exclusion. This problem is investigated in this paper through the design and implementation of a self-organising multi-agent simulator for an iterated cooperative survival game. A series of experiments, or 'survival trials', is run for three different sanctioning schemes: fixed-length, dynamic-length and graduated-length exclusion. Results show that graduated sanctions outperform the alternatives, which can be either too weak or too strong. We conclude by arguing that these results provide evidence for why graduated sanctions are the basis for one of the principles of self-governing institutions.

Keywords: Multi-Agent System · Cooperative Survival Games · Collective Action · Social Contracts · Sanctions · Governance

1 Introduction

Collective action games are a type of game where a player's decisions impact the welfare of the collective as a whole, and all players must work together for a common goal. In these games, there is often an element of cooperative survival, where individuals within the scenario must act in the interest of social welfare, despite their own self interest [18], in order to withstand a disaster. Often, with these games, it is the case that if one player dies, all players die.

These survival games can be considered as a form of extreme, high-stakes common pool of resources (CPR) problem, with the players themselves serving as the common pool. Here, the actions of individuals have consequences for all, such as with the 'tragedy of the commons' [9], where individuals have access to a shared resource that is susceptible to depletion if not properly managed. In this problem, socially oriented traits are necessary to ensure the long-term survival and success of the group. As such, understanding cooperative survival strategies,

and how to incentivise collectivism, is crucial for solving complex problems and achieving sustainability in a range of industries and applications.

For players to successfully traverse these game environments, predicated upon the scarcity of life-sustaining resources, they must embody a number of principles for the management of communities and the common resource in the form of self-governing institutions [19].

This paper investigates the dilemma in developing such an institution in the absence of any central authority. When the only form of legislature is the social construction of social contracts, and the only sanction for non-participation or breaking these contracts is exclusion, then it can be ineffective because exclusion is indistinguishable from non-participation. Moreover, in a high-stakes cooperative survival game, non-cooperation is beneficial in the short-term because the risk of instantaneously dying is eliminated; however it is detrimental to the common good as it increases the probability of the collective dying out sooner than if everyone participated.

The structure of this paper is as follows. We begin by summarising the background of this research in Section 2 by looking at elements of survival games, institutional power, sanctions and social contracts. Following this, we discuss the implementation of the game in Section 3, and the self-organising mechanisms used to solve it in Section 5. Subsequently, we conclude that a simulator is best used to solve this game, which is formalised in Section 6.

In order to examine the effectiveness of sanctioning in such a game, a set of experiments are designed in Section 7 which investigate the survivability of the collective under three different sanction designs: fixed-length, dynamic-length and graduated-length. Here, we conclude that by varying the duration of fixed-length sanctions, a point can be found where survivability is maximised. We also conclude that introducing dynamic and graduated sanctions solves the issue of poor survivability with very low- and high-duration sanction, with graduated sanctions permitting sanctions of effectively maximum length. The simulation platform was developed using GoLang and can be accessed at https://github.com/antonypap/SOMAS2022.

2 Scenario and Background

For the purposes of this paper, we consider a social dilemma where a group of players start at the bottom of a pit, each level of which contains an enemy to be fought. The group must battle and defeat the enemy before they can ascend to the next level, however, any deaths incurred on the way reduce the group's ability to defeat increasingly strong enemies. With each enemy defeated, players get access to a stash of *loot*, containing weapons, shields and potions, which can be divided amongst the group. Weapons are used to attack the enemy, shields are used to defend against the enemy, and potions are used to regenerate health.

Furthermore, this game is designed to be played in an economy of scarcity, meaning that the allocation of loot cannot fully satisfy all of the players' individual desires, leading to biased decision formations, reinforced by increased

individual utility [8] [4]. This condition sets the stage for Ostrom institutions to be formalised for solving a common pool of resources (CPR) game [21], within a norm governed society. These societies take into account the permissions and obligations of its members, as well as the possibility of a deviation from the expected action [1], creating a framework for: sanctions, forgiveness to inspire reconciliation and defiance to incite change [23].

These social norms can be formalised by social contracts, which specify the conditions under which these norms must be obeyed. It has been shown that it is always theoretically possible to design an optimal social contract for the moral imperative [6], although designing this contract is often not a trivial task [24].

As well as defining the conditions by which the social contract must be obeyed, the contract also defines the punishment for not doing so. The breaking of a contract often merits a *sanction* [20], which comes as a detriment to the disobedient actor involved. Such sanctions can vary drastically in severity, such as with their duration, so must be carefully constructed, since "unfair sanctions" [7] can have detrimental impacts on human co-operation. To prevent this, designing effective sanctions has seen a computational approach [2] [17]. In this scenario where sanctions entail exclusion, a negative feedback loop is formed, where sanctioning a defector becomes detrimental to the collective. It is important to prevent free-riders from appropriating the shared resource yet refusing to fight (the risk-averse approach), however over-exclusion will leave them more susceptible to damage, thereby hindering the possibility of co-operative survival. Drawing on the Ancient Greek democratic procedure of Ostracism, which sought to banish tyrannical members of society, damage can be minimised by deposing any unjust institutions who punish defectors with biased sanctions [22].

Social choice theory unifies the relationship between a collection of individual preferences and the final decision of the community [14] [5]. Should these individual preferences be influenced by weighted social knowledge predicated upon reputation, the resulting scenario is an economy of esteem [3], where this reputation is a non-tradeable commodity and cannot be influenced by ones starting position or wealth in a heterogeneous society.

There are various frameworks available to guide decision-making. One such framework is *Preference Utilitarianism*, a contemporary philosophy that seeks to maximize actions that serve the interests of all actors involved [10]. This differs from the conventional "greatest happiness" utilitarian principle [15], as it emphasizes the importance of recognizing the interests of others. In our case, due to an environment of scarcity, it is to be expected that players' personal preference will be to independently accumulate resources. This would entail a lack of recognition of the other, which is essential to morality and ethics [11]. However, despite their condition, we can hope that through knowledge aggregation and collective action, that they can embody preference utilitarianism, by acknowledging that their social network share the same collective interests.

3 Game Design

This game consists of two main phases - a *battle* phase and a *self-organisation* phase - which occur each *level* (l). The battle phase has each player combat the enemy, which runs iteratively until either the enemy is defeated, allowing the players to progress to the next stage, or the players lose (they are killed by the enemy), causing the game to end. If a battle round is victorious, players will progress to the self-organisation phase and subsequently move up a level. The game is completed when the final level is reached (all enemies have been defeated), resulting in a win, or all players have died, resulting in a loss. A compact formalisation of the system architecture is shown in Figure 1, where "S.O. Phase" is an abbreviation for the *self-organisation* phase.

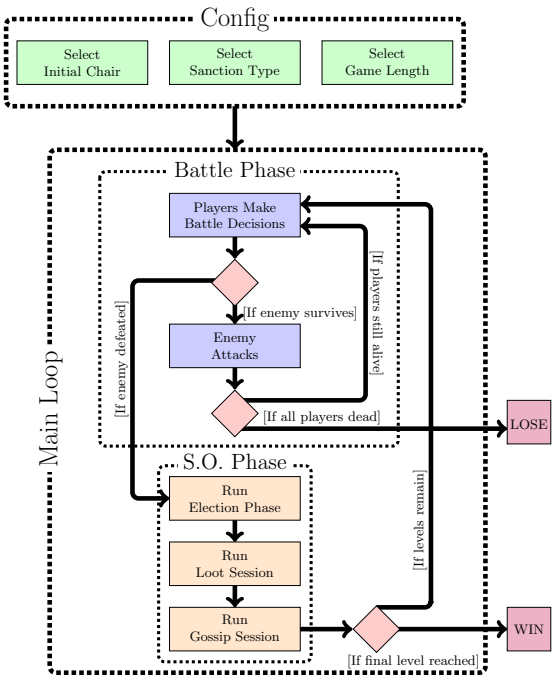


Fig. 1. System architecture for the co-operative survival game

3.1 Entities

We envision three main entities in this game: 1.) the *Players*, 2.) the *Enemy* and 3.) the *Loot*. The *attributes* that these entities possess is shown in Table 1.

These entities are used throughout the game in both the *battle phase* and the *self-organisation* phase. We begin by introducing the *battle phase*, below.

Table 1. List of attributes for each entity

Player	Enemy	Loot
Health (HP)	Resilience (X)	Sword Value (A_s)
Stamina (ST)	Damage Potential (Y)	Shield Value (D_s)
Attack (A)		Health Potion (P_h)
Defence (D)		Stamina Potion (P_s)

4 Battle Phase

In this section, we discuss the different stages in the battle phase, as well as the mathematics behind the enemy.

4.1 Game Stages

Within each battle round, players have three actions that they can perform: *fight*, *defend* and *cower*. Fighting deals damage to the enemy, defending absorbs damage from the enemy and cowering skips the fight round in order to regenerate *stamina* and *health*. *Stamina* is reduced whenever a fight action is performed by the corresponding attribute value, for example, performing a fight action reduces ST by A, so players may only perform an action so long as they have $ST \geq (A \vee D)$. Cowering requires no stamina to perform.

Equipping a loot item affects a player's attributes. *Potions* and *weapons* increase the corresponding attribute by their value. For example: equipping a *sword* increases A by A_s, whilst equipping (drinking) a *health potion* increases HP by P_h. The pit contains a set number of levels (L) and ends if all agents die, or they complete the final level. The rules for battle rounds are as follows:

Rule 1 *An enemy dies if the aggregated attack value of the attacking players is above X at the end of a round.*

Rule 2 *A player dies if the damage received is higher than their remaining HP.*

Rule 3 *If the enemy is not defeated on a given battle turn, it attacks dealing $Y - \sum_i D_i$ (for all defending players) damage divided equally amongst all battling players.*

Rule 4 *If all players cower, they all receive equal damage of $\frac{Y}{N_A}$, where N_A represents the number of players alive on that level.*

4.2 Enemy Formulation

Each player starts with a pre-determined level of health and stamina. Health is only depleted when damage is received from an enemies attack, whilst stamina is depleted with every fight action performed. Both of these attributes are regenerated by their corresponding potion, or through the action of cowering, where agents recover 1% of their starting value.

The enemy attributes: Resilience, X (Equation 1), and Damage Potential, Y (Equation 2), are designed to linearly increase throughout the course of the game. They are also both dependent on: the starting health (HP) and stamina (ST) of the players, the number of players (N), as well as the total number of levels in the pit (L). This helps to maintain the difficulty of the game irrespective of the starting position.

$$X = \delta \frac{N * ST}{L} \sigma \tag{1}$$

$$Y = \delta \frac{N * (HP + ST)}{L} \sigma \tag{2}$$

δ represents a modifier in the range $[0.8, 1.2]$ to add non-determinism to the each separate calculation, σ denotes the linearly increasing scalar (Equation 3).

$$\sigma = \frac{c}{L} + 0.5 \tag{3}$$

where c is the current level. As with the agent fight action values, the damage dealt by the enemy is scaled by a modifier in the range $[0.5, 1]$ applied to the damage potential.

The values given to swords, shields and potions are dependent on the strength of the defeated enemy. Their equations are not included for simplicity. Finally the total quantity of loot dropped is dictated by a pre-determined percentage of N_{init}, the initial number of players in the game, to ensure an economy of scarcity.

5 Self-Organisation Phase

Following the conclusion of a victorious battle round, the game continues with four, successive self-organisation stages:

1. Players may *Gossip* to perform knowledge distribution and aggregation.
2. A vote of No-Confidence is cast to depose the current *Chair* if successful.
3. Elections are held to select a community *Chair*.
4. Players vote on proposed Social Norm contracts to select a new Social Norm.

Participation in these stages is optional for all players, and each stage is discussed in the following sections:

5.1 Gossip, Governance and Social Contracts

Exchanging *gossip* messages is the self-organising mechanism used for knowledge aggregation. This stage of the self-organisation phase allows players to share information about other players by sending a message to a discrete set of recipients. Players then have the ability to update their social perceptions based on this information, however, we consider *false gossip* to be outside of the scope.

The next self-organisational stage allows players to elect a leader, called the *Chair*, who gains *Institutional Powers* as follows: The chair can select and broadcast proposals for voting, and impose sanctions on defectors from the social contract by denying them access to the common pool resource. These powers allow a chair to introduce a bias towards their trusted agents.

Each player has the opportunity to submit themselves for consideration. If elected, their tenure lasts for a maximum of 30 levels, however, at the end of each level, their rule is subject to a no-confidence vote. If a majority is reached, the chair is deposed and a new leader is chosen, this makes leadership strategies a balance between bias and maintaining popularity. Introducing reigns that persist over multiple levels combats the initial transient behaviour within norm governed systems, where the effect of new rulers is not felt of the first few iterations [13].

Social contracts provide a set of mutually agreed conditions under which players must perform certain actions. Each player has the opportunity to create and submit a potential contract, known as a *proposal*, to the elected chair. Each *Battle Contract* contains all four player attributes: HP, ST, A and D, and an associated value for each, as well as a specified action: *attack* or *defend*. This value represents a threshold, with any attribute value over this deeming it 'active'. If all attributes are 'active', a player is obligated to perform the action specified in the contract.

Once a proposal is accepted and the contract is created, players can calculate their required battle action. However, should this action not be in their self-interest, they have the capability of disobeying the contract at the cost of a sanction and being labeled as a 'defector' which may have social implications of a reduction in *reputation*, introduced in Section 6.

5.2 Sanctions

Sanctions, introduced in Section 5.1, deny players access to the common pool resource for a number of levels. Without access, players have no capability of increasing their A or D attributes, as no loot can be obtained. This only leaves agents with the capacity to replenish HP and ST by cowering, an action that could incur further sanctions.

The purpose of these sanctions is to limit 'wasteful' access to the common pool. Players that regularly appropriate from the common pool, however choose to cower, make little use of the items that they obtain. Intuitively, these items would be better given to players that more often comply with the fight contracts, as they will get immediate use out of it. This is especially important given the way that the enemy's damage and health scale according to Equations 1 and 2, as the longer an item is held, the less effective it will be at either dealing or mitigating damage.

The key dilemma with sanctioning is that this exclusion creates a negative feedback loop, as sanctioned players reduce the collective's total potential damage output; it is better to have multiple attackers on a single turn and arm them with shields as well. Having multiple attackers increases the chance of defeating the enemy in a single turn, thus mitigating further damage, and supplying

players with a shield ensures that defending the enemy's attacks is easier. At the same time, these players must be trusted that they will use their items effectively, as cowering will mean that they are effectively wasted.

We propose three different sanctioning mechanisms for affecting the duration that players are excluded from the common pool:

Fixed-Length Sanctions The simplest of methods is a fixed length sanction. In this method, any defectors serve a fixed-length sanction of $l << L$ levels on the interval [0,). We formalise this, and all subsequent sanctions, by introducing the term δl, which represents the change in duration of each successive sanction. Naturally, for fixed-length sanctions:

$$\delta l = 0 \tag{4}$$

Dynamic-Length Sanctions Dynamic sanctions build upon the fixed length method. Chairs are now given the choice of increasing or decreasing the sanction, by a maximum of a single level, depending on the defectors *HP* value. In theory, vulnerable agents would receive smaller sanction severity, thereby increasing the probability of a high average health in the community and increasing the expected utility of a weak player. This method aims to combat the self-defeating feedback loop by showing leniency to weak players.

We consider the *HP* of the collective, HP_c to be normally distributed, and subsequently calculate its mean and variance. This gives the distribution:

$$HP_c \sim \mathcal{N}(\mu_{HP},\, \sigma^2_{HP}) \tag{5}$$

Which influences the change in sanction length, δl, according to Equation 6

$$\delta l = \begin{cases} +1, & HP \geq \mu_{HP} + \sigma_{HP} \\ -1, & HP \leq \mu_{HP} - \sigma_{HP} \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

and ensures that only players at least one standard deviation above or below the mean have varied sanctions. This is in keeping with trying to 'counteract' the negative feedback of sanctions by trying to equalise the accessibility of the common pool.

Graduated-Length Sanctions The final method of determining sanction length is inspired by Ostrom's principles. Graduated sanctions require increasing the sanction length with each repeat defection, up to a maximum sanction level. This method aims to heavily punish repeat offenders, causing them to change to a more collective strategy, whilst mitigating detriment to the collective as a sanction of length 6, say, may better be served in instalments of 1, 2 and 3 successive sanctions. We formalise this sanction in Equation 7

$$\delta l = +1 \tag{7}$$

To tackle the scenario designed in this section, a multi-agent system is adopted, where independent agents act as players within the game.

6 Agent Implementation

In this section, we present the inspiration behind the agent strategy, as well as the primary mechanisms that govern the agent's behaviour. The basis of the agent behaviour is encapsulated in its *reputation*, which is used as the metric by which opinions are formed, trust is built, and sanctions are applied.

Table 2. Parameters used in agent design

Parameter	Range
Reputation (R)	[0,100]
Social Capital (SC)	[0,100]
Trusted Social Network (TSN)	[$Agent$]

In this table are the four principal aspects of the agent design. *Reputation* considers the *needs* and *productivity* of an agent. The evolution of an agent's health and stamina define its *needs*, where a low S and HP imply that the agent is participating in battle (by not cowering), and therefore results in a reputation increase. The agent's decision history over a single level of battle phases is used to determine its *productivity*, where a high fight-to-cower ratio rewards an agent with a reputation increase. Finally, when calculating reputation, an agent considers the opinions of the social network.

Agents send a *gossip* message which consists of a set of (other) agents and their respective reputation scores. Using these *gossip* messages, an agent can perform a weighted average of the information received to update their reputation of the other players. In this scenario, the weights are determined by the communicating agent's level of *trust*.

Social capital rises and falls with the rate of contact between two agents. Continued communication results in an increase in SC, which in turn increases the likelihood of communication in the next gossip session.

Agents can develop opinions about their counterparts using *Trust*, a metric calculated by summing each agent's SC and R, which is then used to determine the player's *Trusted Social Network* (TSN). For each agent the (TSN) is defined as the set of agents with a *trust* score above a certain threshold.

Trust not only influences how agents weight opinions, but also influences the election of the *chair*. The theory that high *reputation* entails high productivity and therefore a level of expertise allows agents to vote for chairs that they believe will be highly effective.

6.1 Preferential Utilitarianism

At the heart of the agent design is a codification of *preference utilitarianism*. To implement this ideology, we use *trust* to split all agents within the environment into two categories: the first being the *TSN*, introduced in Section 6, and the second being the remaining agents. With this network defined, we can summarise this ideology with respect to the cooperative survival game as the choice of actions that maximise the interests of the self, and the *TSN*, whilst endeavouring to satisfice all other agents involved.

Therefore, this agent must be adaptable to the changing circumstances of the social network by creating a dynamically scaling network of agent-to-agent relationships, where actions, disobedience, and peer-to-peer communication influence reputation scores and determine the strength of these relationships.

We envision a direct parallel to Ostrom's perspective on the benefits of non-centralised governance, which in the context of multi-agent systems, may only function in the presence of ubiquitous common knowledge [23]. Therefore, an agent strategy that simultaneously acts to both improve the common knowledge and learn from it, creates a sense of positive feedback.

6.2 Leadership Strategy, Evaluation And Sanctions

When voting on which agent will occupy the position of *Chair*, introduced in Section 5.1, this agent considers a weighted sum of the applicants' *Reputation* and *Social Capital* values. This list of scores is then sorted in descending order to yield a preference order.

The No-Confidence vote mechanism considers social norm disobedience among the *TSN*. An agent measures the performance of the institution by measuring the number of defections in the *TSN*, should this proportion be above a given threshold, the agent votes to depose the chair.

For determining whether to sanction defecting agents, the chair normalises the reputation score of the defecting agent with respect to the reputation of all other agents. If this value is above a given threshold, the agent is sanctioned with the mechanism in use.

6.3 Reputation, Sanctions And Loot Allocation

We introduce the concept *Expected Utility* [16], which is the total amount of utility the agent can produce from an allocated item, to inform loot allocation at the end of the level. As per Section 5.1, agents with a higher reputation are more likely to follow through with fighting, while those with a lower reputation are more likely to renege. Hence, a leader uses the probability $P(U)_i$, of an agent i using an item with a value V_j, to calculate the expected utility $E[V]_i$ of giving agent i the weapon as:

$$E[V]_i = P(U)_i * \sum_j V_j \tag{8}$$

To maximise the expected utility gained by an agent, it is optimal to give the higher valued items to the agents which are more likely to use it (and hence adhere with social contracts). Through using reputation as a naïve indicator of an agent's likelyhood of adherance, the leader sanctions non-compliant agents to maximise the utility of more reliable agents. The non-sancitoned agents are then sorted according to reputation and iteratively given their requested items to ensure the distribution of all loot, as any discarded item has zero effective value. This creates the summation term in Equation 8, where multiple pieces of loot can be allocated to one agent if the number of looting agents is less than the number of items in the loot pool. In reference to Tarantino's *True Romance*, "it's better to have a gun and not need it, than to need a gun and not have it.

6.4 Social Contracts and Collective Actions

Each agent has the opportunity to submit a proposal containing the rules of a potential social contract. According to preferential utilitarianism, this proposal must be designed to maximise the utility of the agent and their *TSN*, whilst ensuring community survival. To achieve this, Equation 9 shows how the threshold for a single agent attribute is calculated, according to the above principals.

$$HP_{Threshold} = HP + (0.2 * \delta_1) + (0.1 * \delta_2) \tag{9}$$

where δ_1 and δ_2 represent the difference between the average collective health and the agent's health, and the average collective health and average *TSN* health respectively. In line with preferential utilitarianism, the weight of personal-to-group state divergence is weighted twice as strongly. This allows the agent to create a proposal resulting in, for example, an attack decision for themselves, whilst modifying the threshold slightly to ensure weak members of the *TSN* are allowed to cower.

7 Experimental Design and Results

With the overarching dilemma of encouraging participation when the only possible sanctioning mechanism is exclusion, we investigate the three categories of sanction introduced in Section 5.2 to establish which method is the most effective for optimising the survivability of the collective.

We assess the survivability by considering the average level reached by the agents, in a simulator comprising 60 levels ($L = 60$), with 30 agents of each type: *Selfless*, *Collective*, and *Selfish*. This yields a total of 90 agents, where all agents are given starting *HP* and *ST* values of 1000 and 2000 respectively.

For each of the test simulations, a parameter sweep of sanction lengths is conducted to produce a line graph showing survivability against sanction length, with each data point averaged across 30 iterations. Each sanction mechanism is simulated with both persistent and non-persistent sanctions to examine the difference between seamless transitions of power and chairs who actively suppress

the decisions of their predecessors. We also note that an additional resilience and potential damage multiplier is applied to X and Y, respectively, to ensure that the game is not trivially winnable across all sanction lengths.

7.1 Fixed Length Sanctions

From the results in Figure 2(a), we see a parabola that peaks at $l = 4$, for the persistent sanctions and $l = 2$ for the non-persistent sanctions. We reason that the trajectory of this figure follows the intuition of the sanctioning mechanism. A 0-length sanction is insufficient in restricting the common pool from free-riders, who will 'waste' the utility of weapons by choosing inaction, resulting in less damage than would otherwise be achieved by prioritising reliable agents and, across multiple turns, less survivability.

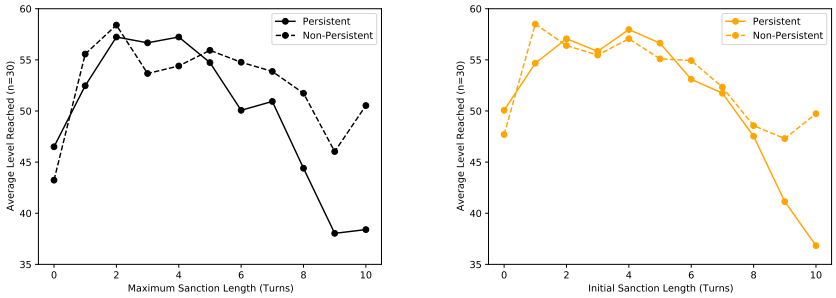
A similar result is found with longer duration sanctions of $l \geq 7$, where the over-exclusion of agents results in equally low survivability in the persistent case. With such long sanctions, it is impossible to effectively arm agents with the swords and shields needed to survive, so the net damage and defence 'potential' of the collective is reduced. Therefore, fewer agents are capable of effectively attacking and/or defending, so defeating the enemy becomes increasingly more challenging. This, again, across multiple turns, results in less survivability.

Intuitively, there is a maximum reached in between these two extremes, where a trade-off between the over-exclusion and under-restriction of the common-pool is achieved. In the persistent case, it is with a sanction duration of $l = 4$, which enables non-compliant agents to be prevented from 'wasting' the high-utility loot, while still enabling them to be sufficiently equipped to remain alive.

A disparity between the persistent and non-persistent sanctions can also be seen in Figure 2(a). We suggest that this is due to the frequency of *Chair* re-elections causing sanctions to effectively be 'forgiven'. For example, a sanction of $l = 7$ may be interrupted after three turns due to a change in *Chair*, resulting in the agent effectively serving an $l = 3$ sanction. It is likely the case that the re-election period is shorter than the sanction duration. This results in the survivability achieved from longer duration, non-persistent sanctions being similar to the survivability of the lower duration, persistent case (a difference of at most eight levels). The rate of survivability decrease is also much slower for non-persistent sanctions.

7.2 Dynamic Length Sanctions

In Figure 2(b), where the x-axis denotes the initial sanction length to which Equation 6 will be applied, a similar parabolic curve to the one described in Section 7.1 can be observed, where an increasing initial duration is followed, $l > 6$, by detrimental effects to the accomplishment of the common goal. Once again, the peak values at $l = 1$ and $l = 4$ dictate the optimal starting points according to this strategy. A noticeable difference can be found on $l = 0$, where there is a dramatic increase in the average level reached. We reason that this is



(a) Fixed sanctions (b) Dynamic sanctions

Fig. 2. Persistent (solid) and non-persistent (dotted), fixed-length (left) and dynamic (right) sanctions

due to the dynamics allowing an increase in sanction length to $l = 1$, resulting in the increase of equipment in the hands of high utility agents.

It is worth noting that both persistent and non-persistent strategies seem to converge to the same results for $2 \leq l \leq 8$. However, the non-persistent sanctions show a global maximum at $l = 1$. A similar divergence between the two approaches can be seen for $l > 8$ as in Figure 2(a), mainly credited to the forgiving nature of sanctions in non-persistent transitions of power.

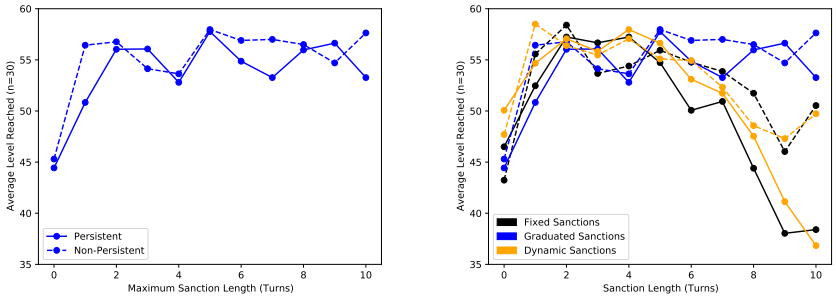
When directly comparing them to the results in Figure 2(a), we can notice that this adaptive approach to sanctioning, modified according to each agents state, shows an improvement on average performance for each sanctioning length. We theorise that this performance increase is due to the increased leniency given to vulnerable agents. Giving these weak agents the opportunity to allocate equipment increases their capabilities, thus increasing the total utility of the collective.

7.3 Graduated Length Sanctions

Following the improvement to the fixed-length sanctions made by the dynamic-length sanctions, we introduce a third and final mechanism of graduated-length sanctions, inspired by Ostrom.

Figure 3(a) deviates from Figures 2(a) and (b) in its trajectory as the sanction length tends to $l = 10$. Here, the survivability trends upwards until a peak at $l = 5$, where it plateaus. This is unlike the previous experiments, where a longer sanction duration was detrimental to the collective.

We reason that this behaviour arises, as agents are never capable of reaching the upper sanction bound of $l \geq 5$, yet are able to effectively serve it in instalments. Reaching an upper bound of, say, $l = 5$ implies that an agent has been sanctioned for a total of ten turns prior to this, effectively serving an $l = 10$ sanction. However, these sanctions are not necessarily served consecutively. Therefore, agents are permitted to access the common-pool to increase their attack and



(a) Graduated sanctions (b) All sanctions overlayed

Fig. 3. Persistent (solid) and non-persistent (dotted), graduated (left) and all (right) sanctions

defence, ensuring that the total utility of the collective increases. This removes the possibility of 'useless' agents, who are incapable of attacking or defending, as it is more likely that every agent has at least one piece of equipment to use.

The lower bound of this plot at $l = 0$ trivially mirrors the behaviour in Figure 2(a), as a maximum length graduated sanction of $l = 0$ is identical to a fixed length sanction of the same duration, with the same issues with survivability.

It is also possible that agents are more incentivised to participate with this sanctioning system, as the plateauing behaviour implies that sanctions of high length are never reached. Graduating sanctions would allow agents to allocate equipment in between sanctions of increasing length, increasing their individual utility whilst also allowing time for an adjusted strategy to take hold.

Ultimately, we see this experiment as unifying the knowledge from experiments reported in Sections 7.1 and 7.2. We have established that a sanction of duration $l = 4$ is effective, however detracting from the collective is detrimental to survivability. Lenient sanctions allow for less-compliant agents to appropriate from the common-pool leading to wasted utility, however it is still important to provide them with the basic means of survival in the event that they may fight in the future, evidenced by the steep decline in survivability in Figures 2(a) and (b). Therefore, we see graduated sanctions as a form of 'trade-off', where a lenient sanction allows for less reliable agents to arm themselves from an early stage, yet be prevented from wasting utility as the game progresses as harsher sanctions are implemented.

7.4 Summary of Experiments

From experiments in Sections 7.1 and 7.2, in Figures 2(a) and (b) we get an inverted parabola for both fixed- and dynamic-length sanctioned. In both cases, an optimum value is reached at $l \approx 4$. Whilst the high-duration performance of both sanction types is similar, the low-duration ($l \leq 2$) is improved for the dynamic-length sanctions.

By introducing graduated-length sanctions in Section 7.3, Figure 3(a) shows how the high-duration ($l \geq 5$) behaviour is improved, resulting in a plateau instead of a decreasing curve. The performance of $l \leq 2$ is unaffected, however, graduated-sanctions with a maximum length of $l <= 2$ are effectively identical to fixed length sanctions of the same duration, due to the under-restriction issue.

Figures 2(a) and (b) show that the disparity between the survivability of persistent and non-persistent sanctions grows as the sanction length increases. This trend is eliminated with graduated-length sanctions, however, in Figure 3(a).

Ultimately, there appears to be three key 'regions' of sanction duration: under-restriction ($l \leq 3$), optimal ($l = 4$) and over-exclusion ($l \geq 5$), which can be seen from Figure 3(b). If the optimal sanction duration is chosen, all methods are equally as effective. If the sanctions are under-restrictive, however, then it is best to chose dynamic-length sanctions and if the sanction are over-exclusive then it is best to choose graduated-length sanctions.

8 Discussion and Further Work

8.1 Discussion

In economics, the Laffer curve has been proposed as showing a theoretical relationship between taxation and revenue [12]. It is argued that with 0% percent taxation, revenue is zero, whilst at 100% taxation, revenue is also zero, as there would be no incentive to work. Therefore, there must be some point in between for the level of taxation which maximises revenue.

By analogy, the same situation appears here: with zero sanctioning, there is no incentive to participate because free-riding is the risk-averse choice; but the ultimate sanction (permanent exclusion) is equally harmful to the collective, since by applying this sanction there will be no one left to participate. It is tempting to postulate that, as with the Laffer curve, there must be some fixed-length sanction duration which maximises the incentive to participate.

However, just as the Laffer curve does not warrant the assertion that cutting taxes increases revenue, starting from a fixed-length sanction and cutting it, as with dynamic-length sanctions, does not solve the problem either. It turns out that graduated sanctions perform best, and there are a variety of reasons for this: including caution (in case of errors and possible appeals); the scope for agents to evaluate opportunity costs, and work out they would be better off participating; and the problem that for *one-shot* wicked problems like common-pool resource sustainability or high-stages cooperative survival, it is simply not possible to run multiple *in vivo* survival trials to find the optimal sanction duration for this particular problem.

We have also seen that, in this type of negative-feedback scenario, it may *not* be effective to have a seamless transition of power between elected chairs. Reneging on the sanctions introduced by one's predecessor may be integral for achieving greater survivability of the collective. This is due to a 'fresh' chair giving offenders a clean-slate, reducing the length of the sanction. Should this period be sufficiently small, it reduces long sanctions to effective levels.

8.2 Further Work

Building on top of the notions of *trust* and *reputation* discussed, we could also explore the nature of posthumous reputation, where agents are conscious of their reputation after their death. This would allow an investigation into any heroic agents, who embody Ambassador Spock's philosophy that "the needs of the many outweigh the needs to the few".

As well as this, we have seen that different sanctions perform well at different lengths. Therefore, a 'mixed-strategy' sanction that combines multiple principles could improve survivability across all sanction lengths. The effect of introducing P2P trading, not restricted by the sanctioning process, could also be explored.

9 Summary and Conclusion

In this paper, we have specified an innovative, co-operative survival game where players are incentivised to participate to maximise collective survival, however the only possible punishment for non-compliance is exclusionary sanctions. This creates a 'negative-feedback loop'. To solve this game, we have developed and specified a self-organising, multi-agent system that facilitates message passing, governance and social contract creation as self-organising mechanisms.

We have investigated three possible techniques for sanctioning non-compliant players: fixed-, dynamic- and graduated-length sanctions, which we assess using a series of survival trial experiments that investigate how the sanction duration for each of the different techniques impacts the survivability of the collective.

We have shown that fixed-length sanctions are feasible, so long as they are carefully tuned to prevent over-exclusion and under-restriction, as the performance is likened to a *Laffer Curve*. We then expand on this by introducing dynamic-length sanctions to offset the negative feedback by increasing and decreasing the sanction length based on the performance of an individual compared to the collective. These help solve the problem of under-restriction, yet still falter in solving the issue of over-exclusion, as the initial duration is too high.

Finally, we unified these two sanction types to implement Ostrom's formulation of graduated-length sanctions by incrementing the fixed-length sanctions by one turn for each successive contract break. This solves the issue of over-exclusion and allows for effectively infinite-length sanctions to be put in place without harming the survivability of the collective. Therefore, we conclude that in a situation where sanctioning is both necessary yet harmful to a collective, implementing graduated-length sanctions is the optimal strategy.

Acknowledgements

We are particularly grateful to the members of the SOMAS team at Imperial College London, specifically, Neel Dugar and Rasvan Rusu for developing a robust infrastructure, as well as Sacha Hakim and Michal Makowka for their contribution to the agent specification.

References

1. Artikis, A., Sergot, M., Pitt, J.: Specifying norm-governed computational societies. ACM Trans. Comput. Logic **10**(1) (jan 2009). https://doi.org/10.1145/1459010.1459011
2. Balke, T., De Vos, M., Padget, J.: I-abm: combining institutional frameworks and agent-based modelling for the design of enforcement policies. Artificial Intelligence and Law **21**(4), 371–398 (Nov 2013). https://doi.org/10.1007/s10506-013-9143-1, https://doi.org/10.1007/s10506-013-9143-1
3. Brennan, G., Pettit, P.: The Economy of Esteem: An Essay on Civil and Political Society. Oxford University Press (03 2004). https://doi.org/10.1093/0199246483.001.0001
4. Cialdini, R.: Influence: The Psychology of Persuasion. NY: William Morrow e Company, New York (1984)
5. Condorcet, N.d.: Essay sur l'Application de l'Analyse à la Probabilité des Décisions Rendue à la Pluralité des Voix. Paris (1785)
6. Davoust, A., Rovatsos, M.: Social contracts for non-cooperative games. In: Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society. p. 43–49. AIES '20, Association for Computing Machinery, New York, NY, USA (2020). https://doi.org/10.1145/3375627.3375829
7. Fehr, E., Rockenbach, B.: Detrimental effects of sanctions on human altruism. Nature **422**(6928), 137–140 (2003)
8. Gigerenzer, G.: How to make cognitive illusions disappear: Beyond "heuristics and biases". European Review of Social Psychology **2**(1), 83–115 (1991). https://doi.org/10.1080/14792779143000033
9. Hardin, G.: The tragedy of the commons. Science **162**(3859), 1243–1248 (1968). https://doi.org/10.1126/science.162.3859.1243, https://www.science.org/doi/abs/10.1126/science.162.3859.1243
10. Hare, R.: Moral Thinking: Its Levels, Method, and Point. Oxford University Press, UK (1981), https://books.google.co.uk/books?id=bp6DdI_f7aQC
11. Hegel, G.W.F.: Phenomenology of Spirit. Oxford University Press, Oxford (1807)
12. Hemming, R., Kay, J.A.: The laffer curve. Fiscal Studies **1**(2), 83–90 (1980), http://www.jstor.org/stable/24434417
13. Kurka, D.B., Pitt, J.: Disobedience as a mechanism of change. In: 12th International Conference on Self-Adaptive and Self-Organizing Systems. IEEE (2018)
14. List, C.: Social Choice Theory. In: Zalta, E.N., Nodelman, U. (eds.) The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University, Winter 2022 edn. (2022)
15. Mill, J.: Utilitarianism. The works of John Stuart Mill, Parker, Son and Bourn (1863), https://books.google.co.uk/books?id=lyUCAAAAQAAJ
16. Mongin, P.: Expected utility theory. Handbook of Economic Methodology pp. 342–350 (01 1997)
17. Nardin, L.G., Balke-Visser, T., Ajmeri, N., Kalia, A.K., Sichman, J.S., Singh, M.P.: Classifying sanctions and designing a conceptual sanctioning process model for socio-technical systems. The Knowledge Engineering Review **31**(2), 142–166 (2016). https://doi.org/10.1017/S0269888916000023
18. Ober, J.: Democracy and Knowledge: Innovation and Learning In Classical Athens. Princeton University Press, Princeton, NJ (2008)
19. Ostrom, E.: Governing the Commons: The Evolution of Institutions for Collective Action. Canto Classics, Cambridge University Press (2015), https://books.google.co.uk/books?id=hHGgCgAAQBAJ

20. Ostrom, E.: Common-pool resources and institutions: Toward a revised theory. Handbook of agricultural economics **2**, 1315–1339 (2002)
21. Ostrom, E.: The challenge of common-pool resources. Environment: Science and Policy for Sustainable Development **50**(4), 8–21 (2008)
22. Perreau de Pinninck, A., Sierra, C., Schorlemmer, M.: Distributed norm enforcement: Ostracism in open multi-agent systems. In: Casanovas, P., Sartor, G., Casellas, N., Rubino, R. (eds.) Computable Models of the Law. pp. 275–290. Springer Berlin Heidelberg, Berlin, Heidelberg (2008). https://doi.org/10.1007/978-3-540-85569-9˙18
23. Pitt, J.: Self-Organising Multi-Agent Systems. World Scientific, London, UK (2021)
24. Scott, M., Dubied, M., Pitt, J.: Social motives and social contracts in cooperative survival games. In: Coordination, Organizations, Institutions, Norms, and Ethics for Governance of Multi-Agent Systems XV: International Workshop, COINE 2022, Virtual Event, May 9, 2022, Revised Selected Papers. pp. 148–166. Springer (2022)