Hybrid Sanctioning for Improved Social Norms Compliance

Yining Yuan[0009−0000−3093−5526], Weiru Liu[0000−0001−8356−1361], and Nirav Ajmeri[0000−0003−3627−097X]

University of Bristol, UK
{yining.yuan, weiru.liu, nirav.ajmeri}@bristol.ac.uk

Abstract. In multi-agent systems (MAS), norms define the expected behaviours, and sanctions enforce them. Most sanctioning models commit to a single enforcement channel, either a centralised authority or decentralised peers. Centralised sanctions are complete but often delayed, e.g., a speeding ticket issued by law enforcement. Whereas decentralised sanctions are often sparse but immediate, e.g., disapproving glances for speaking loudly in a library. We argue that the channel itself should be a design variable. We study hybrid sanctioning as a way to balance these complementary limitations. Our mechanism uses a mixing weight λ to allocate each observed violation to an authority or to an eligible compliant peer. The mechanism does not increase the expected number of sanctions per violation, but redistributes the same enforcement budget. We evaluate this mechanism in a clean-city environment with explicit prohibition and commitment norms, in which agents face a trade-off between efficient travel and proper waste disposal. Across parameter sweeps over population density, hybrid weight, and initial compliance, we find that hybrid sanctioning is an adaptive balance that accounts for the scale of the agent population. We suggest that adapting the sanctioning mechanisms is a productive direction for future work.

Keywords: norm enforcement, sanctioning strategies, agent organisations and institutions

1 Introduction

In MAS, an explicit set of norms, together with the mechanisms that monitor and enforce them, regulates agent behaviour while leaving agents autonomous in whether to comply or violate. Enforcement requires both mechanisms for checking violations and mechanisms for reacting to them through sanctions [17]. Sanctions can be negative reactions triggered by the violation of a norm or positive reactions triggered by compliance with a norm [22]. *Sanctioning* is the process of deciding whether to react, and applying the sanction. *Centralised sanctioning* has a central authority as the sanctioner. *Decentralised sanctioning* is where each agent can be a sanctioner or a sanctionee depending on the state of the world. *Hybrid sanctioning* is a combination of centralised and decentralised sanctioning.

Example 1. Alice and Bob are two agents making trips in a city. They generate rubbish during their trips. The city's cleanliness norm requires rubbish to be placed in public bins. Littering violates this norm. The city employs surveillance cameras to monitor behaviours and issue sanctions at regular intervals. Bob is in a hurry, so he litters. The surveillance camera records Bob's action. A few days later, he receives a fine for violating the cleanliness norm. The fine is the sanction. Because the city issued it, it is centralised sanctioning. Later, Alice throws rubbish on the ground. Bob sees her and knows littering is a violation, so he gives Alice a disapproving look. This disapproving look is a sanction. Because a peer rather than the city issued it, it is decentralised sanctioning. Hybrid sanctioning lets both channels issue sanctions for the same norm.

In the real world, oversight from a centralised authority and decentralised communication coexist [3, 21]. Existing research has largely treated sanctioning as either an institutional mechanism imposed by a central authority or as an emergent peer-to-peer mechanism among autonomous agents. However, each paradigm has structural limitations [33]. Centralised mechanisms depend on designated authorities or norm enforcers who monitor and penalise norm violations, ensuring compliance through a structured, institutionalised framework [15]. Prior work on norm monitoring observes that normative MAS commonly assume that agents' behaviour can be perfectly monitored, while resource-bounded monitoring work explicitly challenges the assumption that monitors have unlimited resources or that monitoring is cost-free [9, 12]. While these authorities can enhance completeness, they often entail high maintenance costs [5, 9, 12], and the rewards for compliance are often delayed because authority observation and adjudication operate at coarser timescales than the agents being regulated [3]. In contrast, decentralised sanctioning mechanisms operate in open, distributed agent societies where individual agents apply punishment based on their observations of norm violations [26, 29].

In both paradigms, the cost of sanctioning falls on whoever occupies the sanctioner role; what differs is who that is and how the cost is borne. Under centralised sanctioning, the cost is an institutional overhead, absorbed by the authority and amortised across the population. Under decentralised sanctioning, it falls directly on an ordinary agent that has its own goals to pursue, so the cost is deducted from that agent's task performance and can be substantial [16], especially for non-altruistic agents. Decentralised agents also have only local, partial perceptions of the environment, which constrains what the system can infer about compliance [2] and leads to incomplete evaluations. Effectiveness is further contingent on norm salience [20] and on the second-order free-riding problem, i.e., agents benefiting from enforcement without bearing its cost [4]. Sparse populations yield fewer observation events, weakening the signal that drives compliance learning. These complementary limitations suggest hybrid regimes in which centralised and decentralised enforcement operate simultaneously; their interaction remains under-explored [33]. The key design problem is how to allocate enforcement responsibility between an institutional channel that is broad but temporally limited and a peer channel that is immediate but spatially limited.

Prior work in evolutionary game theory shows that centralised and peer punishment can coexist and interact non-linearly [28]. A follow-up combination happens in a collective-risk social dilemma in which cooperators report defection while a monitoring and sanctioning institution imposes fines [18]. Recent MARL work examines how multiple social enforcement mechanisms, such as direct punishment, third-party punishment, reputation, and partner selection, interact during learning [13]. However, this line of work focuses on peer-level social mechanisms rather than the coexistence of institutional and peer sanctioning channels. We study hybrid sanction in normative MAS where agents act in sequential environments and norms and sanctions are explicitly represented.

Contributions Specifically, our central research question is: *How does the balance between centralised and decentralised sanctioning affect norm compliance, environmental outcomes, and sanction efficiency in norm-aware agent societies?*

To investigate our central question, we create Hyena (*hybrid enforcement of norms in agent societies*), a hybrid sanctioning mechanism that combines centralised and decentralised enforcement as complementary channels with different temporal and spatial observability: the institutional channel has broad monitoring capacity but enforces only at discrete intervals, whereas the peer channel has only local observability but reacts immediately. We evaluate Hyena in a clean-city simulation based on Example 1, sweeping population density, enforcement timing, hybrid mixing weight, initial compliance, and compliance-learning sensitivity. We find that the effectiveness of hybrid sanctioning is conditional on interaction density: it improves compliance when peer observations are frequent enough to complement centralised enforcement, whereas in sparse settings reducing centralised enforcement weakens coverage.

Organisation Section 2 summarises the related work. Section 3 formalises the components of agents, norms and sanctions. Section 4 describes our model of a clean-city environment and its parameters. Section 5 reports the comparison between hybrid sanctioning and the two single-channel baselines. Section 6 presents conclusions and discussions.

2 Related Works

Research on sanctioning mechanisms for norm compliance and adoption is related to our contributions. A complementary line of work studies how the norms themselves should change over time. Savarimuthu et al. [26] study how agents can dynamically infer, add, modify, and delete conditional norms by observing interactions without explicit communication, using a park littering scenario. We create a similar littering scenario grounded by their work. Dell'Anna et al. [14] introduce a centralised norm-revision component that dynamically modifies sanction magnitudes from observed violation rates and estimates agent preferences using Bayesian Networks to keep non-compliance within acceptable bounds. Prior work on norm emergence and norm learning asks how agents infer or revise norms.

We study a setting in which the norm is already specified and available to agents, but compliance remains an open decision because agents have task incentives that may conflict with it. Our representation of norms and sanctions follows Singh's [27] model.

Sanctions can take various forms. A sanction can combine material punishment and norm communication [3], or be operationalised as constraints in which violations remain possible but costly and informative [34]. Peer reputation also drives compliance pressure [10]. Considering the cost of enforcement, the marginal utility of sanctions suggests that only a certain proportion of enforcement agents increases cooperation [5]. Sanctions are also studied as an adaptive mechanism tied to group performance: adaptive sanctioning strategies enhance cooperation in a public goods game [21] by empowering agents to decide which sanction to apply based on factors such as violation severity, the offender's history, or context. This focus on informed decision-making aligns with our approach.

Centralised norm enforcement in mixed-motive MARL means enhancing the environment with a normative system controlled by an external RL regulator [11]. Santos et al. [25] propose a centralised framework to classify actions and provide violators with feedback on the features that caused the violation. However, centralised coordination becomes impractical in distributed MAS. Yan et al. [33] propose an agent-centric approach for BDI agents, using the Normative Programming Language to represent centralised norms while maintaining each agent's autonomy to detect norm violations. A decentralised setting raises the question of where the sanctioning criterion comes from when no authority publishes one. One answer is to learn it: classifiers categorise behaviours as socially approved or disapproved, so peers acquire a shared standard from observation rather than from a specification [7, 29], and training sanction actions on spurious norms can generalise sanctioning behaviour to new norms in multi-agent reinforcement learning [20]. Our norms and sanctions are specified rather than learnt, so what the agent learns is not the criterion but the expectation of how reliably that criterion is enforced.

3 Method

In this section, we formalise our agent structure, environment and sanctions.

3.1 Formalisation

We first fix the environment, since it introduces the sets that the remaining definitions refer to.

Definition 1 (Environment). *An environment is a tuple* $\langle \mathbb{A}, \mathcal{S}, \mathcal{N}, F \rangle$*, where* $\mathbb{A}$ *is the set of all agents,* $\mathcal{S}$ *is the set of states,* $\mathcal{N}$ *is the set of all norms, and* F *is the set of sanction transition functions related to each norm.*

We write Expr for the set of logical expressions over $\mathcal{S}$, each of which evaluates to true or false in a given state $s \in \mathcal{S}$, and $s \models \varphi$ for "φ holds in s".

An agent is an autonomous entity that perceives the environment, makes decisions, and takes actions to achieve its goals.

Definition 2 (Agent). *An agent $ag \in \mathbb{A}$ is a tuple $\langle G, O, A, C\rangle$ where G is its set of goals; O is its observation space, A is its set of possible actions, where $a \in A$ is a single action; and $C = [c_1, \ldots, c_{|\mathcal{N}|}]$ is the agent's norm-compliance attitude vector, where $c_n \in [0, 1]$ is agent's compliance probability with norm $n \in \mathcal{N}$.*

For instance, an agent in our Example 1 can be represented as $ag\langle\{\text{reachGoal}\}, O, \{\text{move}, \text{drop}, \text{pickUp}, \text{sanction}\}, [c_{n_1}, c_{n_2}]\rangle$, where O is the set of observations available within its observation zone, and $[c_{n_1}, c_{n_2}]$ is its compliance attitude towards the two norms.

A norm governs the agents' interaction in an environment. We consider two types of norms following [27]: *commitments* (Cmm) and *prohibitions* (Pro).

Definition 3 (Norm). *A norm is a tuple n: $\langle m, \textsc{sbj}^n, \textsc{obj}^n, \mathsf{ant}^n, \mathsf{con}^n\rangle$, where the norm type $m \in \{Cmm, Pro\}$; $\textsc{sbj} \subseteq \mathbb{A}$ is its subject, the set of agents who need to comply; $\textsc{obj} \subseteq \mathbb{A}$ is its object, the set of agents who need to enforce; $\mathsf{ant} \in$ Expr is its antecedent; and $\mathsf{con} \in$ Expr is its consequent. Here Expr denotes the set of logic expressions over the state space $\mathcal{S}$ that are either true or false.*

A commitment $\langle Cmm, \text{SBJ}, \text{OBJ}, \mathsf{ant}, \mathsf{con}\rangle$ states that SBJ commits to OBJ to bring about con whenever ant holds. For instance, the norm of disposing rubbish in the bin is represented as n_1: $\langle Cmm$ Alice, pedestrian, hasRubbish, dispose$\rangle$. A prohibition $\langle Pro, \text{SBJ}, \text{OBJ}, \mathsf{ant}, \mathsf{con}\rangle$ states that, accountable to OBJ, SBJ must not bring about con whenever ant holds. For instance, the norm of not littering is represented as n_2: $\langle Pro$, Alice, city, hasRubbish, litter$\rangle$.

Norm compliance means that the agent's action satisfies the norm. Following the commitment lifecycle [19], a commitment is satisfied when, after ant holds, SBJ's actions bring about con, and is violated otherwise. A prohibition is complied with when, while ant holds, con does not occur, and is violated when con occurs.

Definition 4 (Norm Compliance). *Let $s \in \mathcal{S}$ be the state in which a is taken and let $s \xrightarrow{a} s'$ denote the resulting state. The transition (s, a, s') complies norm n, written $(s, a, s') \Vdash n$, when:*

$$(s, a, s') \Vdash n \iff s \not\models \mathsf{ant} \vee \begin{cases} s' \models \mathsf{con} & \text{if } m = Cmm \\ s' \not\models \mathsf{con} & \text{if } m = Pro \end{cases} \tag{1}$$

Compliance is then evaluated per agent:

$$comp(ag, n, (s, a, s')) = \begin{cases} True & \text{if } ag \notin \textsc{sbj}^n \text{ or } (s, a, s') \Vdash n \\ False & \text{otherwise} \end{cases} \tag{2}$$

An agent outside $\textsc{sbj}^n$ is not bound by n and so cannot violate it.

Sanctions are reactions to norm compliance or violation. A sanction can be positive (triggered by compliance) or negative (triggered by violation). In this work, we consider only negative sanctions. We map definitions of commitment and prohibition norms to sanctions as defined by Singh [27]. We distinguish between the specification of a sanction in Definition 5 and its transition function in Equation 4.

Definition 5 (Sanction). *A sanction σ is a tuple $\langle n, \textsc{sbj}^\sigma, \textsc{obj}^\sigma, \mathsf{ant}^\sigma, \mathsf{con}^\sigma \rangle$ where n is the associated norm, $\textsc{sbj}^\sigma \subseteq \mathbb{A}$ is the sanctionee set, $\textsc{obj}^\sigma \subseteq \mathbb{A}$ is the sanctioner set, $\mathsf{ant}^\sigma \in$ Expr is the condition that triggers the sanction, and con^σ is the consequent imposed on $\textsc{sbj}^\sigma$, which could be a reward, transition function, a deterministic sequence of actions, or a combination of both.*

For an agent in OBJ, whether to sanction an agent in OBJ is a strategic action, not a deterministic reaction. Even if an agent $ag_i \in \textsc{obj}^\sigma$ has observed a violation of n by some other agent $ag_j \in \textsc{sbj}^\sigma$, ag_i must still decide whether to apply the consequent. This decision may depend on the cost of sanctioning, on ag_i's goals G, and on its own compliance attitude C, all as given in Definition 2. Therefore, for $\exists ag_j \in \textsc{sbj}^\sigma, ag_i \in \textsc{obj}^\sigma$:

$$\mathsf{ant}^\sigma \equiv (\neg comp(ag_j, n, a_j)) \wedge (a_i = sanction) \tag{3}$$

The conjunct $a_i = sanction$ is a condition on the sanctioner's chosen action is to sanction at this step.

To describe runtime behaviour, we define the sanction transition function $sanc \in F$, which maps a sanction and the sanctioner's action to the consequent to be applied:

$$sanc(\sigma, a_i) = \begin{cases} \mathsf{con}^\sigma & \text{if } s \models \mathsf{ant}^\sigma, \\ \bot & \text{otherwise} \end{cases} \tag{4}$$

For the negative sanctions we consider, these state updates and forced actions specified in con^σ are goal-impeding by increasing costs, delaying attainment, or complicating the attainment of the sanctionee's goals.

One component of con^σ is bookkeeping. Let ς_n^t denote the number of sanctions an agent has received for violating norm n up to time t, recorded in the sanctionee's internal state. Applying σ increments it, $\varsigma_n^t \leftarrow \varsigma_n^t + 1$. It is used in updating compliance in Section 3.3. The cost borne by the sanctioner is real but of a different kind, i.e., time and forgone progress rather than a normative record, and we account for it separately in the sanctions-per-agent metric of Section 4.5.

3.2 Sanction Mechanisms

We present three sanction mechanisms — centralised, decentralised and hybrid — that differ only in who occupies the role of sanctioner.

In centralised sanctioning, only an institutional agent observes violations with full observability.

Definition 6 (Centralised Sanctioning). *The sanctioner is only an institutional agent* $OBJ^{\sigma} = ag^{\star}, SBJ^{\sigma} = \mathbb{A}$. *The sanctioner's policy is deterministic:*

$$a_* = sanction \text{ if } \neg comp(ag_j, n, a_j) \wedge (ag_* \text{ observes } ag_j) \tag{5}$$

In decentralised sanctioning, peer agents only partially observe violations within their observation zone O_t^{ag}.

Definition 7 (Decentralised Sanctioning). *Every agent is a potential enforcer of every other agent:* $OBJ^{\sigma} = \mathbb{A}, SBJ^{\sigma} = \mathbb{A}$. *Each* ag_i *chooses* $a_i = sanction$ *strategically in its own policy.*

$$\mathsf{ant}^{\sigma} \equiv \neg comp(ag_j, n, a_j) \wedge (ag_i \text{ observes } ag_j) \wedge (a_i = sanction) \tag{6}$$

Each peer ag_i observes only what falls within its observation zone $O_t^{ag_i}$ at the current step. A violation is detectable by a peer only at the moment it occurs and only by those nearby. Once the violator has moved on, the evidence is no longer within any peer's observation window. The institutional sanctioner ag^* has access to the full system state and can observe every violation. However, real-world institutional enforcement typically does not act on every individual violation in real time. It batches evidence at the end of a decision epoch, after a fixed interval, on receipt of a complaint, etc.

Definition 8 (Hybrid Sanctioning). *The sanctioner set is the union of both channels,* $OBJ^{\sigma} = \mathbb{A} \cup ag^{\star}, SBJ^{\sigma} = \mathbb{A}$ *with stochastic role assignment controlled by a sanction allocation parameter* $\lambda \in [0, 1]$.

Upon observing a violation, a peer sanction is issued with probability λ, and a centralised sanction with probability $1 - \lambda$. Setting $\lambda = 1$ recovers the decentralised mechanism, $\lambda = 0$ recovers the centralised mechanism. Intermediate values mix the two, so a single violation may be enforced by a peer, by the institution, by both, or by neither in a given run.

3.3 Compliance Update

Situated artificial institutions show that applying norms in MAS requires linking abstract institutional concepts to concrete environmental states and events [8]. We assume each agent has access to the specifications of n and σ, which means the agent can reason about the expressions $\mathsf{ant}^n, \mathsf{con}^n, \mathsf{ant}^{\sigma}$, and con^{σ} against its observation O_t^{ag}.

The norm-compliance attitude is widely used in agent-based modelling for simulating littering and cleaning behaviour [24]. The norm-compliance attitude c_n^{ag} in each agent records the extent to which the agent complies with each norm. At the start of each decision epoch, agent ag samples a compliance intention $intend_n^{ag}$ with respect to each norm $n \in \mathcal{N}$. The intention is a latent realisation of the agent's c_n^{ag}, drawn once per epoch and held fixed during this epoch:

$$intend_n^{ag} = \begin{cases} comply & \text{if } \xi_n < c_n^{ag}, \\ violate & \text{otherwise,} \end{cases} \qquad \xi_n \sim \mathcal{U}(0,1). \tag{7}$$

The intention constrains the agent's action policy over the epoch. Under $intent_n^{ag} = comply$, every action a_t^{ag} taken during the epoch satisfies $comp(ag, n, a_t^{ag}) = \text{True}$. Under $intent_n^{ag} = violate$, the agent's policy admits at least one a_t^{ag} with $comp(ag, n, a_t^{ag}) = \text{False}$.

During the decision epoch, agents treat sanctions as evidence that the norm is enforced and adjust their strategy accordingly. Each compliance attitude is updated based on sanctions in Definition 5. Compliance attitude improves with sanctions, but a forgetting term must also be considered.

For each agent ag and norm n, the agent maintains an empirical expectation $e_n^t \in [0,1]$ that the norm is socially shared, and a normative expectation $m_n^t \in [0,1]$ that the rule is centrally enforced [6].

We model that the compliance attitude is derived from the two beliefs,

$$c_n^t = \text{clip}_{[0,1]}(w_e e_n^t + w_m m_n^t + w_{\text{int}} e_n^t m_n^t), \tag{8}$$

with weights $w_e, w_m, w_{\text{int}} \geq 0$ representing the intrinsic importance of the empirical expectation, the normative expectation and their combination. Let $Q_n^t \in \{0,1\}$ indicate that ag received at least one peer-sanction event inside its observation radius, $K_n^t \in \{0,1\}$ indicate that it received a centralised sanction, and U_n^t is the number of unsanctioned violations. $e_n^{t+1} = e_n^t + \alpha_e Q_n^t (1 - e_n^t) - \beta U_n^t e_n^t$, and $m_n^{t+1} = m_n^t + \alpha_m K^t (1 - m_n^t) - \beta U_n^t m_n^t$. The rates $\alpha_e, \alpha_m, \beta \in [0,1]$ represent the strength of positive evidence from receiving peer sanctions and centralised sanctions, and the forgetting induced by an unpunished violation. The compliance attitude c_n^t is recomputed using Equation 8 at each decision step.

The learning rates $\alpha_e, \alpha_m, \beta$ and the weights w_e, w_m, w_{int} are fixed exogenous parameters, so that the comparison across λ is not confounded by simultaneously tuned learning dynamics. We set $w_e = w_m$ and vary w_{int} to isolate the effect of the interaction term.

4 Experiment

To evaluate the hybrid sanctioning mechanism, we develop a simulated clean-city environment using Python and Pygame [23]. Agents in the simulation use heuristic-based planning to update their personal objectives and normative behaviours. **Reproducibility.** Codebase is available at [35].

4.1 Clean-City Environment

The clean-city environment is modelled as a grid world as illustrated in Figure 1. $\mathcal{D} = [W] \times [H]$ is a finite 2-D grid. This grid consists of locations for public spaces $P = \{P_1, \ldots, P_n\}$ and for designated waste-disposal bins $B = \{B_1, \ldots, B_n\}$.

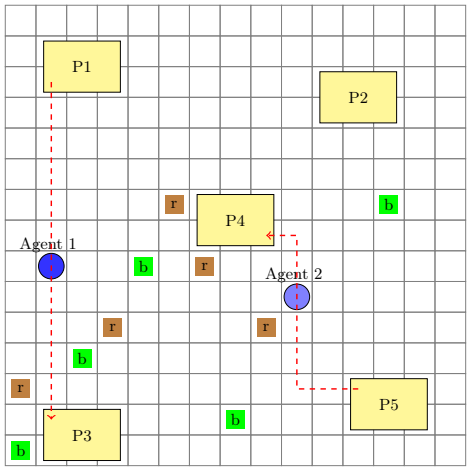


Fig. 1: The clean-city environment has a grid layout. Here, random cells are set as public places (P1–P5). A cell with 'b' represents a bin. A cell with 'r' represents a cell with rubbish. Dashed lines represent agents' paths.

$R_t \subseteq \mathcal{D}$ is the set of cells currently containing litter. P and B are reset each episode, and R is updated every step.

A state $s_t \in \mathcal{S}$ consists of $s_t = (\mathbf{x}_t,\ \mathbf{h}_t,\ \boldsymbol{\varsigma}_t,\ \mathbf{c}_t,\ R_t)$. For each agent ag, $x_t^{ag} \in \mathcal{D}$ is the agent's location, $h_t^{ag} \in \{1, 0\}$ represents if the agent has rubbish or not, $\varsigma_t^{ag} \in \mathbb{N}$ is the cumulative sanction count, $c_t^{ag} \in [0, 1]$ is the compliance attitude indicating the extent to which they are compliant to the norms of the society. The destination $g_t^{ag} \in P$ and planned path π^{ag} are the agent's internal variables refreshed at each trip start.

Agent ag's observation $o_t^{ag} = \Omega(s_t, ag)$ contains $\{x_t^j, h_t^j, \varsigma_t^j\}$, where j represents neighbour agents that $j : |x_t^j - x_t^{ag}| \leq 1$, littering events occurring in t within the same neighbourhood, and internal states $\{x_t^{ag}, h_t^{ag}, \varsigma_t^{ag}, c_t^{ag}\}$. In the centralised society, each agent, except ag^*, has the action set $A = \{move, wait, drop, pickup\}$. ag^* only has the action $A = \{sanction\}$. In the decentralised society, each agent has the action set $A = \{move, wait, drop, pickup, sanction\}$. In the hybrid society, each agent excluding ag^* has $A = \{move, wait, drop, pickup, sanction\}$. ag^* only has the action $A = \{sanction\}$. Agents' transitions after each action are deterministic, unless two or more agents try to move to the same cell. Movement conflicts are resolved by randomly assigning one agent to move and the other agents to remain in place. When $a_t = drop$, rubbish is added to their current position $R_{t+1}^x = R_t^x + 1$. When an agent disposes of rubbish in a bin, their rubbish-carrying status resets $h^{ag} = 0$. When an agent performs action $a_t =$ sanction targeting another agent j who littered, the target's sanction count $\varsigma_{t+1}^j = \varsigma_t^j + 1$.

4.2 Norms and Sanctions

We consider two norms in the society, a commitment and a prohibition. They are operationally distinct but are two facets of the same normative situation. $n_1 : \langle Cmm, \textsc{citizens}, ag^*, (h_t^{ag} = 1), dispose \rangle$, which reads as: A citizen is committed to the city to throw rubbish in a bin. $n_2 : \langle Pro, \textsc{citizens}, ag^*, (h_t^{ag} = 1 \wedge x_t^{ag} \notin B), drop \rangle$, which reads as: a citizen is prohibited by the city from throwing rubbish on the ground. The condition $x_t^{ag} \notin B$ restricts the prohibition to dropping outside a public space, since a public space is where a bin is reachable and dropping there is not the littering the norm targets. Dropping while outside a public space is littering.

The same sanction σ is attached to both n_1 and n_2. If an agent is sanctioned for littering, it picks up the nearest litter and takes it to the nearest bin. It is represented as $\sigma : \langle n, \textsc{citizens}, \textsc{obj}^\sigma, \mathsf{ant}^\sigma, \mathsf{con}^\sigma \rangle$. Here, $\textsc{obj}^\sigma$ depends on the mechanism:

$$\textsc{obj}^\sigma = \begin{cases} ag^* & \text{centralised,} \\ \textsc{citizens} & \text{decentralised,} \\ \textsc{citizens} \cup \{ag^\star\} & \text{hybrid.} \end{cases}$$

To model the institutional cost without introducing a full authority-resource model, we approximate limited institutional capacity by allowing the authority to evaluate ant^σ only at fixed intervals of τ steps before issuing sanctions. We restrict peer sanctioning to agents who were themselves norm-compliant on the current trip. The antecedent ant^σ specialises Equation 3 to this environment: $\mathsf{ant}^\sigma \equiv (a_{t-1}^j = drop \wedge x_{t-1}^j \notin B) \wedge (j \in O_t^{ag_i}) \wedge (a_t^i = sanction)$. The consequent con^σ has two components: $\varsigma_{t+1}^j \leftarrow \varsigma_t^j + 1$ and $\pi^j \leftarrow \textsc{PathToLitterThenBin}(x_t^j, R_t, B)$. The first component records sanction sources and times, which drives the compliance update of c^j in Equation 8 at the end of the epoch. The PathToLitterThenBin overrides the sanctioned agent's plan: it must first *pickup* litter at the nearest cell in R_t, then route via the nearest bin in B to *drop*, then resume its original path. The forced detour increases the trip length, which represents the goal-impeding consequent in the sense of Definition 5.

4.3 Agent's Decision Process

The agent's goal is to reach the goal state as fast as possible while incurring as little sanction cost as possible. Agents navigate public spaces, generating rubbish at the start of each trip. The cost of compliance is the additional time required to travel to the bins. The cost of sanctioning for citizen agents is waiting one timestep for the sanctioned agent to pick up the litter, and the cost for the authority agent is measured by the number of sanctions it issues.

At the start of each trip, an agent ag samples a destination $g^{ag} \sim \text{Uniform}(P \setminus x_t^{ag})$ as the goal state excluding its current cell, and draws its compliance intention $intend_n^{ag}$ based on Equation 7. The intention selects one of two planning modes, both implemented by Breadth-First Search (BFS):

- **Norm-compliant routing** ($intend_n^{ag} = comply$) : the agent computes a BFS path that is constrained to pass through the nearest bin before reaching the destination, modelled as a shortest-path problem, represented as $x^{ag}t \to B_m \to g^{ag}$, where $B_m = \arg\min B \in \mathcal{B}|x_t^{ag} - B|_1$;
- **Direct routing** ($intend_n^{ag} = violate$) : the agent computes a single-leg BFS path directly to g^{ag}. If it is carrying rubbish $h_t^{ag} = 1$, the first litter-free cell outside a public space on this path triggers a deterministic *drop* action.

Then, the agent set can be partitioned into $\mathbb{A}_t^C = \{ag \in \mathbb{A} : intend_n^{ag} = comply\}$ and $\mathbb{A}_t^V = \{ag \in \mathbb{A} : intend_n^{ag} = violate\}$.

The deterministic *drop* action represents the urge to dispose of the trash as quickly as possible. BFS is used because other agents dynamically occupy cells and act as soft obstacles that require re-planning on conflict, and the bin-via constraint in compliant routing is a heuristic search over a two-stage goal graph. When multiple shortest paths exist, we break ties in favour of the path whose cells have the smallest total Manhattan distance to the nearest bin, i.e., the route that stays closest to a bin, so that a compliant agent's detour is minimal.

At each step t, the agent executes the next action on its plan and observes O_t^{ag}. If $ag \in \text{OBJ}^\sigma$ and $intent_n^{ag} = comply$, it may additionally select $a_t^{ag} = sanction$ for an observed violator j, subject to the antecedent in Equation 3. Incoming sanctions are recorded in ς_t^{ag}, and any forced action sequence from con^σ overwrites the agent's own plan until completed. When an agent reaches g^{ag}, the epoch ends. It updates its compliance attitude and randomly picks another public place as a new destination. The experiment terminates when the maximum run length is reached, or the highest compliance is reached.

4.4 Environment Parameters

We test the hybrid mechanism on grids of size $|[W] \times [H]| = 20 \times 20$ and the density of society using parameters in Table 1. We run each setting for 10 random initial maps to mitigate stochastic effects.

4.5 Metrics

We compute metrics for compliance, environmental outcomes, and cost of sanctions. For compliance, $\mathbf{M_{CP}}$ is the mean compliance attitude of the population at each time step, $\mathbf{M_{CP}} = \frac{1}{|\mathbb{A}|}\sum_{ag\in\mathbb{A}} c_{n,t}^{ag}$. $\mathbf{M_{AC}}$ is the proportion of compliant agents in the population at each step, $\mathbf{M_{AC}} = \frac{|\mathbb{A}_t^C|}{|\mathbb{A}|}$.

For environmental outcomes, $\mathbf{M_{CL}}$ is the percentage of clean cells at each time step, $\mathbf{M_{CL}} = \frac{|\mathcal{D}\setminus R_t|}{|\mathcal{D}|}$; $\mathbf{M_G}$ is the average of the goals achieved in each run.

For the cost of sanctions, $\mathbf{M_S}$ is the cumulative number of sanctions issued in the population up to step t, $\mathbf{M_S} = \sum_{ag\in\mathbb{A}} \varsigma_t^{ag}$; $\mathbf{M_U}$ is the cumulative number of unsanctioned violations in the population up to step t, $\mathbf{M_U} = \sum_{ag\in\mathbb{A}} u_t^{ag}$; $\mathbf{M_{UP}} = \frac{\mathbf{M_U}}{(\mathbf{M_U}+\mathbf{M_S})}$ is the unsanctioned proportion. Sanctions per agent $\mathbf{M_{SP}}$ measures the average sanction issued by each agent.

Table 1: Environment and simulation parameters. Bins, public places, and agents scale with density; τ is set as a multiple of the per-condition trip length. Values in braces are swept. Single-valued entries are held fixed across all runs.

Variable	Description	Sparse	Natural	Dense
$\lvert\mathbb{A}\rvert$	Number of agents	12	40	133
$\lvert P\rvert, \lvert B\rvert$	Number of public places, bins	12	20	40
T	Run length	1000	2000	5000
c_n^0	Initial compliance attitude	{0, 0.5, 1}		
w_e	Importance of empirical expectation	{0.5}		
w_m	Importance of normative expectation	{0.5}		
w_{int}	Importance of their combination	{0, 0.4}		
α_e	Strength of peer sanction	{0.2}		
α_m	Strength of centralised sanction	{0.2}		
β	Strength of unsanctioned violation	{0.1}		
λ	Hybrid mixing parameter	{0, 0.25, 0.5, 0.75, 1}		
τ	Central-check interval	{2}		

5 Results

We evaluate sanctioning mechanisms, fully centralised ($\lambda = 0$), fully decentralised ($\lambda = 1$) and the proposed hybrid allocation ($\lambda \in \{0.25, 0.5, 0.75\}$) across three population densities.

5.1 Compliant Behaviour

Compliance converges to substantially different equilibria across sanctioning mechanisms, as shown in Figure 2 and Table 2. In the dense world, the hybrid sanctioning drives the highest final mean compliance $\mathbf{M_{CP}} = 0.95$. In the natural world, hybrid again leads at $\mathbf{M_{CP}} = 0.79$. In the sparse world, the three mechanisms are statistically indistinguishable on $\mathbf{M_{CP}}$. Compliance collapses under fully decentralised peer sanctioning at every density, whereas hybrid sanctioning matches or exceeds centralised sanctioning wherever peer observations are frequent enough to contribute, and it never falls below it.

5.2 Environmental Outcomes

Figure 3 shows the boxplots of goals reached across parameters, and Figure 4 shows the boxplots of the final cleanliness reached across parameters. In sparse and natural worlds, $\mathbf{M_G}$ is insensitive to λ, whereas in the dense society, the hybrid sanctioning at $\lambda = 0.5$ and $\lambda = 0.75$ reaches the fewest goal states. In sparse and natural worlds, $\mathbf{M_{CL}}$ is close to the environment bound, whereas in the dense society, hybrid sanctioning at $\lambda = 0.5$ and $\lambda = 0.75$ results in a cleaner environment. This occurs because peer-augmented enforcement converts

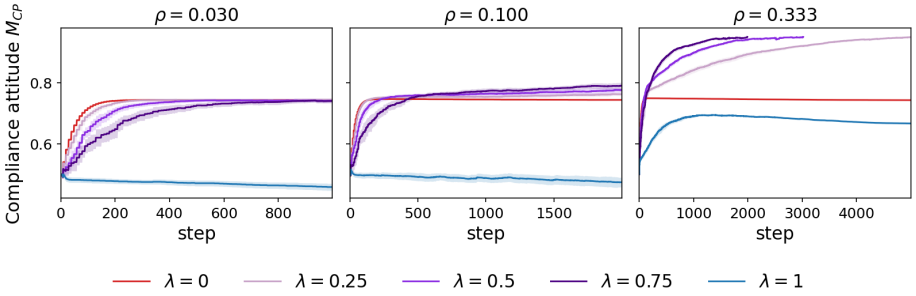


Fig. 2: Compliance attitude $\mathbf{M_{CP}}$ over time, mean across runs with 95% CI band, by world and density. ρ represents the density of the population. $\rho = 0.030$ is sparse, $\rho = 0.100$ is natural and $\rho = 0.333$ is dense.

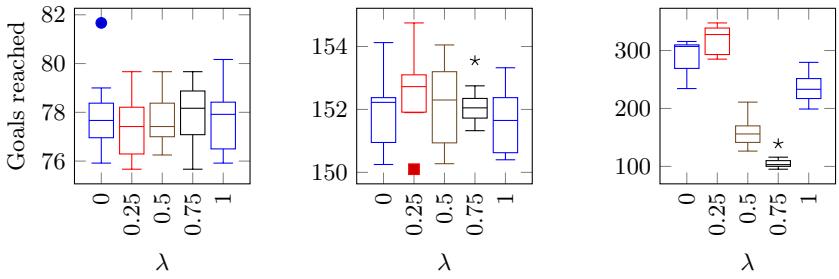
Table 2: Metrics across densities (mean±95% CI across seeds).

Density	λ	$\mathbf{M_{CP}}$	$\mathbf{M_{AC}}$	$\mathbf{M_{CL}}$	$\mathbf{M_G}$	$\mathbf{M_{UP}}$	$\mathbf{M_{SP}}$
Dense	0	0.74±0.00	0.64±0.01	49.40±1.52	293.49±17.05	0.03±0.00	8.90±0.56
	0.5	0.95±0.00	0.77±0.01	67.63±1.93	196.66±33.81	0.14±0.01	5.35±0.53
	1	0.67±0.00	0.57±0.01	49.45±2.04	238.94±16.11	0.11±0.01	27.61±2.92
Natural	0	0.74±0.00	0.57±0.02	88.42±0.54	152.03±0.71	0.10±0.01	2.33±0.07
	0.5	0.78±0.01	0.58±0.01	88.97±0.29	152.37±0.38	0.36±0.05	1.70±0.14
	1	0.47±0.02	0.36±0.02	86.22±0.74	151.74±0.61	0.69±0.04	1.33±0.17
Sparse	0	0.74±0.00	0.25±0.06	97.33±0.36	78.05±0.97	0.20±0.05	0.91±0.11
	0.5	0.74±0.00	0.30±0.03	97.27±0.17	77.84±0.42	0.44±0.10	0.63±0.13
	1	0.46±0.01	0.15±0.04	96.88±0.33	77.90±0.80	0.89±0.05	0.12±0.06

attitudinal compliance into actual cleaning moves, as specified in the sanction's consequent.

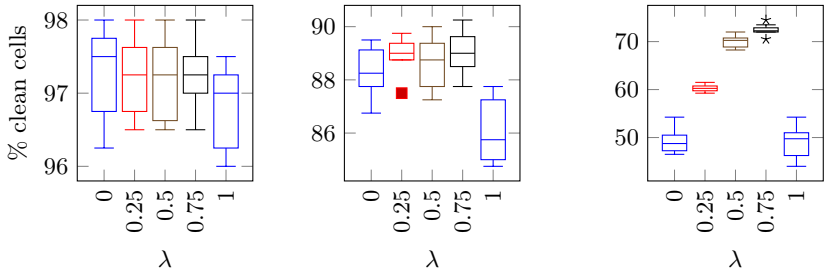
5.3 Sanction Cost

Table 2 also lists the goals reached $\mathbf{M_G}$, the unsanctioned proportion $\mathbf{M_{UP}}$ and the sanctions given per agent $\mathbf{M_{SP}}$. The compliance gains reported above are obtained at very different enforcement costs, and the cost structure itself changes with density, so we read the table density-by-density. In the dense world, the cost gap is largest: decentralised agents issue an average of 27.6 sanctions each, roughly five times the 5.35 of hybrid at $\lambda = 0.5$, while centralised sits at 8.90. Here, hybrid buys the highest compliance at the lowest per-agent enforcement load. In the natural world, the decentralised is the cheapest. However, decentralised leaves 69% of violations unsanctioned compared to 0.10 for centralised, so its low count reflects enforcement that mostly does not occur rather than efficient enforcement. The sparse world sharpens the same effect, leaving 89% of violations unsanctioned, which is why its compliance collapses. Goals reached $\mathbf{M_G}$ are essentially flat across mechanisms at natural and sparse density.



(a) Sparse. (b) Natural. (c) Dense.

Fig. 3: Final goals reached $\mathbf{M_G}$ with λ in densities of the world. λ=0 is centralised, λ=1 is decentralised; intermediate values are hybrid.



(a) Sparse. (b) Natural. (c) Dense.

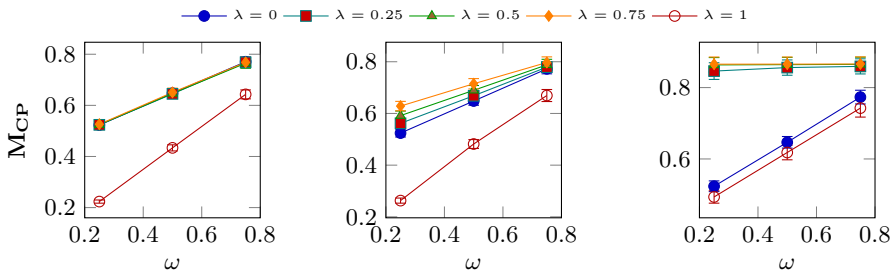
Fig. 4: Final clean cells with λ in densities of the world. λ=0 is centralised, λ=1 is decentralised; intermediate values are hybrid.

5.4 Sensitivity Studies

Instead of having all agents $c_n^0 = 0.5$, we initialise the proportion of compliant agents and non-compliant agents. ω is proportion of the agent population starts with $c = 1$ and $1 - \omega$ starts with $c = 0$. $\omega = \{0.25, 0.5, 0.75\}$. $\mathbf{M_{CP}}$ under hybrid is essentially flat in ω (0.864, 0.864, 0.865 for $\omega = 0.25, 0.5, 0.75$), while pure centralised and pure decentralised rise monotonically with ω.

6 Conclusion and Discussion

We study hybrid sanctioning, which combines centralised and decentralised enforcement as complementary channels. Our findings demonstrate that the hybrid sanctioning generally outperforms both decentralised and centralised sanctioning mechanisms in enhancing average norm compliance among agents across various environments. This effectiveness is particularly pronounced in densely populated environments, where interactions between agents are frequent. However, in sparse environments, the hybrid sanctioning mechanism is less



(a) Sparse (b) Natural (c) Dense

Fig. 5: Final $\mathbf{M_{CP}}$ across $\omega \in \{0.25, 0.5, 0.75\}$ under $\lambda \in \{0, 0.25, 0.5, 0.75, 1\}$.

effective than the centralised method. This is primarily due to the limited opportunities for interaction between agents, as reducing centralised enforcement through a hybrid sanction weight can diminish overall enforcement effectiveness.

While our experiments provide several key insights, some threats to validity warrant consideration. First, the system has unavoidable randomness for agent goals as well as parameters. We mitigate the threat arising from randomness by conducting multiple iterations of the experiments. Second, due to limited access to real behavioural datasets, the parameters in the compliance update function and our hybrid weights are hypothetical and fixed. However, these numbers are empirically selected. Further, to mitigate this threat, we conduct experiments on a variety of society types.

Directions Our analysis is based on one littering scenario. One direction is to evaluate hybrid sanctioning in other cooperative and mixed-motive social scenarios to determine how the relative effectiveness of enforcement mechanisms varies across contexts. Another direction is to study norm internalisation by withdrawing sanctions after a period of enforcement. Such experiments would help distinguish compliance sustained by enforcement from internalised compliance. Extending the model to represent agents' values could further inform mechanisms to guide the broader sociotechnical system toward value-norm equilibrium [1].

Our current simulation simplifies agents' decision-making, which may be influenced by multiple, potentially conflicting norms. Future work could examine how agents reason about such conflicts and decide whether to comply despite the costs involved. Rather than treating all violations and agents uniformly, the type and intensity of sanctions could be guided by ethical principles, for example, by protecting the worst-off while balancing social welfare [30, 32].

The current mechanism imposes a sanction using a hard-coded probability. A more realistic mechanism would account for the effort required to sanction and its effect on an agent's willingness to enforce a norm. Future work could explore adaptive sanction allocation that accounts for these costs while remaining transparent and explainable [31].

Bibliography

[1] Ajmeri, N., Vos, M.D., Dell'Anna, D., Murukannaiah, P.K., Nallur, V., Nardin, L.G., Singh, M.P.: Guiding sociotechnical systems toward value-norm equilibrium. In: Proc. 25th International Conference on Autonomous Agents and Multiagent Systems. pp. 3869–3875. IFAAMAS, Paphos (May 2026). https://doi.org/10.65109/NSZQ3158, Blue Sky Ideas Track

[2] Alechina, N., Halpern, J.Y., Kash, I.A., Logan, B.: Incentivising monitoring in open normative systems. In: Proc. 31st AAAI Conference on Artificial Intelligence. pp. 305–311. AAAI Press, San Francisco (2017). https://doi.org/10.1609/aaai.v31i1.10610

[3] Andrighetto, G., Brandts, J., Conte, R., Sabater-Mir, J., Solaz, H., Villatoro, D.: Punish and voice: Punishment enhances cooperation when combined with norm-signalling. PLOS ONE **8**(6), e64941 (2013). https://doi.org/10.1371/journal.pone.0064941

[4] Axelrod, R.: An evolutionary approach to norms. American Political Science Review **80**(4), 1095–1111 (Dec 1986). https://doi.org/10.2307/1960858

[5] Balke, T., De Vos, M., Padget, J.: Evaluating the cost of enforcement by agent-based simulation: A wireless mobile grid example. In: Proc. 16th Principles and Practice of Multi-Agent Systems. LNCS, vol. 8291, pp. 21–36. Springer, Dunedin (2013). https://doi.org/10.1007/978-3-642-44927-7_3

[6] Bicchieri, C., Dimant, E., Gächter, S., Nosenzo, D.: Social proximity and the erosion of norm compliance. Games and Economic Behavior **132**, 59–72 (Mar 2022). https://doi.org/10.1016/j.geb.2021.11.012

[7] Boyd, R., Mathew, S.: Arbitration supports reciprocity when there are frequent perception errors. Nature Human Behaviour **5**(5), 596–603 (May 2021). https://doi.org/10.1038/s41562-020-01008-1

[8] de Brito, M., Hübner, J.F., Boissier, O.: Situated artificial institutions: stability, consistency, and flexibility in the regulation of agent societies. Autonomous Agents and Multi-Agent Systems **32**(2), 219–251 (Mar 2018). https://doi.org/10.1007/s10458-017-9379-3

[9] Bulling, N., Dastani, M., Knobbout, M.: Monitoring norm violations in multi-agent systems. In: Proc. 12th International Conference on Autonomous Agents and Multiagent Systems. pp. 491–498. IFAAMAS, St. Paul (2013). https://doi.org/10.5555/2484920.2484999

[10] Castelfranchi, C., Conte, R., Paolucci, M.: Normative reputation and the costs of compliance. Journal of Artificial Societies and Social Simulation **1**(3) (1998), https://www.jasss.org/1/3/3.html

[11] Cheang, R.M., Brandão, A.A.F., Sichman, J.S.: Centralized norm enforcement in mixed-motive multiagent reinforcement learning. In: Proc. International Workshop on Coordination, Organizations, Institutions, Norms, and Ethics for Governance of Multi-Agent Systems. pp. 121–133. LNCS, Springer (2022). https://doi.org/10.1007/978-3-031-20845-4_8

[12] Criado, N.: Resource-bounded norm monitoring in multi-agent systems. Journal of Artificial Intelligence Research **62**(1), 153–192 (May 2018). https://doi.org/10.1613/jair.1.11206

[13] Dasgupta, N., Musolesi, M.: Investigating the impact of direct punishment on the emergence of cooperation in multi-agent reinforcement learning systems. Autonomous Agents and Multi-Agent Systems **39**(1), 19 (Mar 2025). https://doi.org/10.1007/s10458-025-09698-5

[14] Dell'Anna, D., Dastani, M., Dalpiaz, F.: Runtime revision of sanctions in normative multi-agent systems. Autonomous Agents and Multi-Agent Systems **34**(2), 43 (2020). https://doi.org/10.1007/s10458-020-09465-8

[15] Esteva, M., Rosell, B., Rodríguez-Aguilar, J.A., Arcos, J.L.: Ameli: An agent-based middleware for electronic institutions. In: Proc. 3rd International Joint Conference on Autonomous Agents and Multiagent Systems. pp. 236–243. IEEE, New York (2004). https://doi.org/10.1109/AAMAS.2004.10060

[16] Fehr, E., Gächter, S.: Altruistic punishment in humans. Nature **415**(6868), 137–140 (2002). https://doi.org/10.1038/415137a

[17] Grossi, D., Aldewereld, H., Dignum, F.: Ubi Lex, Ibi Poena: Designing norm enforcement in e-institutions. In: Proc. International Workshop on Coordination, Organizations, Institutions, and Norms in Agent Systems. pp. 101–114. LNCS, Springer, Hakodate (2007). https://doi.org/10.1007/978-3-540-74459-7_7

[18] He, N., Chen, X., Szolnoki, A.: Central governance based on monitoring and reporting solves the collective-risk social dilemma. Applied Mathematics and Computation **347**, 334–341 (2019). https://doi.org/10.1016/j.amc.2018.11.029

[19] Kafali, Ö., Ajmeri, N., Singh, M.P.: Desen: Specification of sociotechnical systems via patterns of regulation and control. ACM Transactions on Software Engineering and Methodology **29**(1), 7:1–7:50 (Jan 2020). https://doi.org/10.1145/3365664

[20] Köster, R., Hadfield-Menell, D., Everett, R., Weidinger, L., Hadfield, G.K., Leibo, J.Z.: Spurious normativity enhances learning of compliance and enforcement behavior in artificial agents. Proc. National Academy of Sciences **119**(3), e2106028118 (2022). https://doi.org/10.1073/pnas.2106028118

[21] de Lima, I.C.A., Nardin, L.G., Sichman, J.S.: Gavel: A sanctioning enforcement framework. In: Proc. 6th International Workshop on Engineering Multi-Agent Systems. LNCS, vol. 11375, pp. 225–241. Springer, Stockholm (2019). https://doi.org/10.1007/978-3-030-25693-7_12

[22] 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 (Mar 2016). https://doi.org/10.1017/S0269888916000023

[23] Pygame Community: Pygame (2026), https://www.pygame.org, accessed 2026-07-15

[24] Rangoni, R., Jager, W.: Social dynamics of littering and adaptive cleaning strategies explored using agent-based modelling. Journal of Artificial Societies

and Social Simulation **20**(2), 1 (2017). https://doi.org/10.18564/jasss.3269

[25] Freitas dos Santos, T., Osman, N., Schorlemmer, M.: Learning for detecting norm violation in online communities. In: Proc. International Workshop on Coordination, Organizations, Institutions, Norms, and Ethics for Governance of Multi-Agent Systems. pp. 127–142. LNCS, Springer, London (2022). https://doi.org/10.1007/978-3-031-16617-4_9

[26] Savarimuthu, B.T.R., Cranefield, S., Purvis, M.A., Purvis, M.K.: Identifying conditional norms in multi-agent societies. In: Proc. International Workshop on Coordination, Organizations, Institutions, and Norms in Agent Systems (COIN). pp. 285–302. LNCS, Springer, Berlin (2011). https://doi.org/10.1007/978-3-642-21268-0_16

[27] Singh, M.P.: Norms as a basis for governing sociotechnical systems. ACM Transactions on Intelligent Systems and Technology **5**(1), 21:1–21:23 (Dec 2013). https://doi.org/10.1145/2542182.2542203

[28] Szolnoki, A., Szabó, G., Czakó, L.: Competition of individual and institutional punishments in spatial public goods games. Physical Review E **84**, 046106 (Oct 2011). https://doi.org/10.1103/PhysRevE.84.046106

[29] Vinitsky, E., Köster, R., Agapiou, J.P., Duéñez-Guzmán, E.A., Vezhnevets, A.S., Leibo, J.Z.: A learning agent that acquires social norms from public sanctions in decentralized multi-agent settings. Collective Intelligence **2**(2), 1–14 (Apr 2023). https://doi.org/10.1177/26339137231162025

[30] Woodgate, J., Ajmeri, N.: Macro ethics principles for responsible AI systems: Taxonomy and directions. ACM Computing Surveys **56**(11), 1–37 (Jul 2024). https://doi.org/10.1145/3672394

[31] Woodgate, J., Collins, D.E., Yuan, Y., Ajmeri, N.: Reframing transparency, interpretability, and explainability for multi-agent systems. In: Proc. International Workshop on Coordination, Organizations, Institutions, Norms and Ethics for Governance of Multi-Agent Systems. pp. 1–34. Paphos (2026)

[32] Woodgate, J., Marshall, P., Ajmeri, N.: Operationalising rawlsian ethics for fairness in norm-learning agents. In: Proc. 39th AAAI Conference on Artificial Intelligence. pp. 2382–26390. AAAI, Philadelphia (Feb 2025). https://doi.org/10.1609/aaai.v39i25.34837

[33] Yan, E., Nardin, L.G., Hübner, J.F., Boissier, O.: An agent-centric perspective on norm enforcement and sanctions. In: Proc. International Workshop on Coordination, Organizations, Institutions, Norms, and Ethics for Governance of Multi-Agent Systems. pp. 79–99. LNCS, Springer, Auckland (2025). https://doi.org/10.1007/978-3-031-82039-7_6

[34] Yan, E., Nardin, L.G., Boissier, O., Sichman, J.S.: A unified view on regulation management in multi-agent systems. In: Proc. International Workshop on Coordination, Organizations, Institutions, Norms, and Ethics for Governance of Multi-Agent Systems. pp. 55–74. LNCS, Springer, Detroit (2026). https://doi.org/10.1007/978-3-032-17542-7_4

[35] Yuan, Y.: Hybrid sanctioning for improved social norms compliance. Zenodo (2026). https://doi.org/10.5281/zenodo.21380255, accessed 2026-07-15