Runtime Revision of Norms and Sanctions based on Agent Preferences

Davide Dell’Anna
Utrecht University
Utrecht, The Netherlands
d.dellanna@uu.nl

Mehdi Dastani
Utrecht University
Utrecht, The Netherlands
m.m.dastani@uu.nl

Fabiano Dalpiaz
Utrecht University
Utrecht, The Netherlands
f.dalpiaz@uu.nl

ABSTRACT

To fulfill the overall objectives of a multiagent system, the behavior of individual agents should be controlled and coordinated. Runtime norm enforcement is one way to do so without over-constraining the agents’ autonomy. Due to the dynamicity and uncertainty of the environment, however, it is hard to specify norms that, when enforced, will fulfill the system-level objectives in every operating context. In this paper, we propose a mechanism for the automated revision of norms by altering their sanctions, based on the data monitored during the system execution and on some knowledge about the agents’ preferences. We use a Bayesian Network to learn at runtime the relationship between the obedience/violation of a norm and the achievement of the system objectives. We propose two heuristic strategies that explore the updated Bayesian Network and automatically revise the sanction of an enforced norm. An evaluation of our heuristics using a traffic simulator shows that our mechanisms outperform uninformed heuristics in terms of convergence speed.

KEYWORDS

Multiagent Systems; Norm Revision; Norm Enforcement

ACM Reference Format:

1 INTRODUCTION

A multiagent system (MAS) comprises a set of autonomous agents that interact in a shared environment [20]. To reach the emergent system-level objectives of a MAS, the behavior of the autonomous agents should be coordinated [5]. For example, a smart traffic system is a MAS that includes autonomous agents like cars, traffic lights, etc. The objectives of such a system include ensuring that each agent reaches its destination timely, and that the number of accidents is minimized.

Norm enforcement is a prominent mechanism for controlling and coordinating the runtime behavior of the agents in a MAS without over-constraining their autonomy [1, 18]. Norm enforcement via sanctions is traditionally contrasted with norm regimentation, which prevents the agents from reaching certain states of affairs. For example, in a smart traffic system, a regimentation strategy could be to close a road to prevent cars from entering the road, while a sanctioning strategy could be to impose sanctions on cars that do enter the road.

Existing research has studied the revision of the enforced norms, and proposed logics that support norm change [4, 13, 14], introduced algorithms for switching among alternative norms [10], and studied the legal effects of norm change [11]. Dell’Anna et al. [10] have proposed a framework for engineering normative MASs that, using runtime data from MAS execution, revises the norms in the MAS to maximize the achievement of the system objectives. Their work makes the simplistic assumption that norms are regimented.

In this paper, we make a step forward toward the engineering of normative multiagent systems that can revise norms enforced via sanctions. Moreover, and in contrast to [10], norm revision is now informed about the preferences of the agents. We assume preferences to be specified in terms of a desired state of affairs and the maximum payment that the agent is willing to make to achieve the state of affairs. We use Bayesian Networks to learn the norm effectiveness from runtime execution data and to inform the norm revision mechanism that revises the sanction of the enforced norm.

The contributions of this paper is as follows:

- We introduce a simple formal framework to specify a normative multiagent system in terms of norms and agent preferences (agent types).
- We build on and extend the general architecture proposed in [10], and study the relationships between preferences, sanctions, and system objectives.
- We propose heuristics for norm revision that make use of probabilistic information learned from system execution data to align the enacted norms with the agent preferences.
- We report on an evaluation through a traffic simulator that shows the effectiveness of our revision strategies in identifying optimal sanctions efficiently.

Organization. Section 2 presents our framework to characterize norms and agent preferences. Section 3 explains the overall approach for supervising normative MAS based on probabilistic reasoning over norms’ effectiveness. Section 4 introduces our strategies for revising norms by combining agent preferences with the satisfaction of the overall system objectives. Section 5 evaluates our work through simulation experiments. Section 6 discusses related work, and Section 7 presents our conclusions and future directions.

2 NORMATIVE MULTIAGENT SYSTEMS

This section presents a simple, but generic, framework for specifying normative multiagent systems in which the agents behave according to their preferences while norms are enforced on them.
via sanctions. This framework allows us to analyze the interplay between norms and preferences in normative multiagent systems.

2.1 Illustrative example

We consider the two-lanes ring road depicted in Figure 1. In a ring road, a population of vehicles moves continuously in circle. Every vehicle is autonomous and acts according to its own preferences (e.g., it respects a certain speed limit and is willing to risk certain sanctions). If a fast vehicle is using the outer line and a slower vehicle blocks its way, the fast vehicle moves to the inner line to overtake the slow vehicle. Since all vehicles share the same environment, their local decisions have an effect on the emergent system-level behavior of the ring road [17]. For example, based on contextual factors such as the density of vehicles in the ring road, the vehicles behavior may provoke traffic jams and the average speed may vary, as well as the average time to complete a round of the ring road.

Consider a set of propositional atoms \( \left\{ a_1, \ldots, a_n \right\} \), each representing a state of affairs. A norm is a pair \( N = (p, s) \), where \( p \in L \) and \( s \in \mathbb{N} \), indicating that \( p \) should hold in the current system state, otherwise sanction \( s \) will incur. For instance, given \( L = \left\{ \text{speed}_50, \text{speed}_100 \right\} \), a norm \( N = \left( \text{speed}_50, 100 \right) \) indicates that every vehicle in the ring road shall not exceed a speed of 50km/h, otherwise it will receive a sanction of 100€.

In this paper we focus on rational agents, i.e., agents that always choose to achieve their most preferred state of affairs. Let \( Aq = \{ a_1, \ldots, a_n \} \) be a set of agents. The preference of an agent \( a \) is denoted by \( \text{Pref}(a) = (C, \geq) \), where \( C = \left\{ (p_i, b_i) \mid 1 \leq i \leq k \land b_i \in \mathbb{N} \right\} \) and \( \geq \) is a partial order on \( C \). A pair \( (p, b) \in \text{Pref}(a) \) indicates that the agent \( a \) wants to achieve \( p \) and is willing to spend a budget \( b \), i.e., the agent is willing to pay \( b \) or less to achieve \( p \). Given \( L \) as defined above and restricting us to two possible budget values \( 0 \) and \( 1 \), the preference \( (\text{speed}_50, 0) \geq (\text{speed}_100, 1) \geq (\text{speed}_50, 0) \) indicates that the agent prefers to drive fast rather than slow, and that maximizing speed has priority over minimizing the budget. As an agent’s preference characterizes its type, this agent can be seen as having a brave type.

An agent is said to have a basic rational preference \((C, \geq)\) when for all pairs \( (p_i, b_i) \) and \( (p_j, b_j) \in C \), the partial order \( \geq \) satisfies one of the following two conditions:

1. \((p_i, b_i) \geq (p_j, b_j) \) iff \( b_i \leq b_j \) & \( \forall (p, b), (p', b') \in C, \forall b, b' \in \mathbb{N} : (p, b) \geq (p', b') \Rightarrow (p, b') \geq (p', b') \).

2. \((p_i, b_i) \geq (p_j, b_j) \) iff \( (p_i = p_j) \) then \( b_i \leq b_j \) else \( \forall b, b' \in \mathbb{N} : (p_i, b) \geq (p_j, b') \).

If an agent’s preference is defined according to the first clause, then the agent’s budget is used to order the pairs. If the preference adheres to the second clause, then the propositional atom is used to order the pairs. An example of non-rational preference, instead, is \((\text{speed}_50, 1) \geq (\text{speed}_50, 0) \geq \ldots \), since the first two pairs share the same propositional atom but, among them, it is preferred the pair with higher budget. Without going into the details, which are outside the scope of this paper, note that an agent’s preference as defined above satisfies the rationality requirements, i.e., the basic rational preference is transitive and complete.

We also consider more complex agent preferences that combine the two clauses of basic rational preference. An agent is said to have a complex rational preference, or simply a rational preference \((C_1 \cup C_2, \geq) \) iff \((C_1, \geq_1) \) and \((C_2, \geq_2) \) are basic agent preferences, \( C_1 \cap C_2 = \emptyset \), the budgets that are used in \( C_1 \) are lower than any budget used in \( C_2 \), and \( \geq \geq_1 \cup \geq_2 \cup \left\{ (c_1, c_2) \mid c_1 \in C_1 \land c_2 \in C_2 \right\} \). Again, without going into the details that go beyond this paper, note that the rational preference is transitive and complete.

Given a norm \( N = (p, s) \) and an agent preference \((C, \geq)\) with \((p', b) \in C\), we assume that (i) \( p, p' \in L \), and (ii) \( s, b \in \mathbb{N} \). In other words, we assume that norms and agents’ preferences are defined with respect to the same environment and that the sanction and the agents’ budget are commensurable. This makes it possible to analyze an agent’s preference in the context of a norm to determine whether the preference can motivate an agent to comply with a norm or otherwise to which extent it will violate the norm.

Intuitively, in the context of an enforced norm, we assume that an agent follows its preference by aiming to realize either a norm complying state or a violating state for which he is willing to pay the sanction.
assume that a preferred state \( p_1 \) (e.g., \( \text{speed}_{-100} \)) violates a norm state \( p \) (e.g., \( \text{speed}_{-50} \)) when \( p \) excludes \( p_1 \), and we write \( \text{viol}(p_1, p) \) to indicate it. In the sequel, the pair \((p_1, b)\) is said to be a violating pair w.r.t. a norm \((p, s)\) when \( \text{viol}(p_1, p) \); otherwise \((p_1, b)\) is said to be a complying pair w.r.t. a norm \((p, s)\).

Let \((C, \geq)\) be the preference of an agent and \( N = (p, s) \) be a norm. A pair \( c_i = (p_i, b_i) \in C \) is called the most preferred (state/budget) pair to act upon in the context of \( N \) if and only if for all pairs \( c_j = (p_j, b_j) \in C \) with \( c_j \geq c_i \) it holds that \( \text{viol}(p_j, p) \) and \( b < s \). For example, given the agent preference \((\text{speed}_{-100}, 0) \geq (\text{speed}_{-100}, 1) \geq (\text{speed}_{-50}, 0) \geq (\text{speed}_{-50}, 1) \) and the norm \((\text{speed}_{-50}, 2)\), the most preferred pair to act upon is \((\text{speed}_{-50}, 0)\). Note that given a norm \((p, s)\) it is possible to have the most preferred pair \((p_i, b_i)\) such that \( \text{viol}(p_i, p) \), i.e., the most preferred pair motivates the agent to realize a violating state. For example, for the above-mentioned agent preference, and a norm \((\text{speed}_{-50}, 1)\) that states “do not exceed 50km/h to avoid sanction 1”, the most preferred pair to act upon is \((\text{speed}_{-100}, 1)\). This pair indicates that the agent is willing to pay sanction 1 for having a speed of 100km/h. We say that an agent \( a \) has a reason to violate a norm \((p, s)\) whenever the agent’s preference \( \text{Pref}(a) \) is such that the most preferred pair is \((p_i, b_i)\) and \( \text{viol}(p_i, p) \).

In order to determine the maximal payment that an agent is willing to pay for violating a given norm, we introduce the notion of maximum budget for norm violation. Let \( \text{Pref}(a) = (C, \geq) \) be the preference of agent \( a \) and \( N = (p, s) \) be the enforced norm. The maximum budget that \( a \) is willing to pay for the violation of \( N \), denoted as \( \text{maxB}(a, N) \), is the highest budget of the violating pairs in \( C \) that are preferred over the most preferred complying pair in \( C \), i.e., the highest budget that occurs in the set: \( \{ (p_i, b_i) \in C \mid (p_i, b_i) \geq (p, b) \in C \} \), with \((p, b)\) the first complying pair in \( \text{Pref}(a) \). For example, consider \((\text{speed}_{-100}, 0) \geq (\text{speed}_{-100}, 1) \geq (\text{speed}_{-50}, 0) \geq (\text{speed}_{-50}, 1) \) as the preference of agent \( a \) and \( N = (\text{speed}_{-50}, 2) \) as the enforced norm. The maximum budget that \( a \) is willing to pay for violating \( N \) is 1, i.e., \( \text{maxB}(a, N) = 1 \). Note that if the maximum budget for violating a norm is lower than the sanction corresponding with the norm, then the most preferred pair to act upon is a complying pair.

### 3 NORM-BASED SUPERVISION

We build on the runtime norm-based supervision systems for multiagent systems as proposed in [10] and sketched in Figure 2. The framework monitors the behavior of a multiagent system, evaluates the enforcement of the norms in terms of the overall system objectives, and, when needed, intervenes by revising the norms.

In such a framework, a Bayesian Network called Norm Bayesian Network is used to learn and reason at runtime about the correlation between norm obedience or violation and the satisfaction of the system-level objectives. For example, the approach helps answering questions like how well, and in which contexts, does the norm \((\text{speed}_{-100}, 2)\) help achieve the objective of avoiding halted cars?

In [10] we proposed heuristic algorithms for suggesting norm revisions that alter the regimented norms. Here, we use the framework to revise the way the norms are enforced by altering the sanctions. We adopt their Norm Bayesian Network as a tool to identify correlations between norms and system-level objectives.

#### 3.1 Norm Bayesian Network

Consider some monitorable environmental properties such as the density of vehicles or the presence of an obstacle in the ring road. Each of these properties is called contextual variable, and is associated to a domain of values. For example, Vehicles density can be either low or high, while Obstacle can be true or false. Given a set of contextual variables, a context assigns a value to each contextual variable. For instance, given Vehicles density and Obstacle, four possible contexts exist: high-true, high-false, low-true, low-false.

A Norm Bayesian Network \( NBN = (X, A, P) \) [10] is a Bayesian Network where:

- \( X = N \cup O \cup C \) are nodes that represent random variables in probability theory. \( N, O \) and \( C \) are disjoint sets. \( N \) consists of norm nodes; each node \( N \in N \) corresponds to a norm and has a discrete domain of 3 possible values: obeyed, violated and disabled. \( O \) consists of objective nodes; each node \( O \in O \) corresponds to a boolean objective and has a discrete domain of 2 values: true and false. Finally, \( C \) consists of context nodes; each node \( C \in C \) corresponds to a contextual variable and can have a discrete or continuous domain of values.
- \( A \subseteq (C \times N) \cup (C \times O) \cup (N \times O) \) is the set of arrows that connect pairs of nodes. If there is an arrow from node \( X \) to node \( Y \), \( X \) is called parent of \( Y \).
- \( P \) is a set of conditional probability distributions. These are encoded into conditional probability tables (CPTs), each one associated with a node in \( X \) and quantifying the effect of the parents on the node. The conditional probability values in the CPT of a node are the parameters of the network. These parameters are learned from runtime monitoring data through classic Bayesian learning.

In the rest of the paper, we use the following notation for Bayesian Networks. Italic uppercase (\( X, \ldots \)) for random variables; bold uppercase (\( X, Y, \ldots \)) for sets of random variables; italic lowercase (\( v_1, v_2, \ldots \)) for values in the domain of a random variable; \( N_v \) abbreviates \( N = v \), i.e., an assignment of value \( v \) to a norm variable \( N \); \( O_v \) denotes an assignment of value \( v \) to all nodes in \( O \); \( P \) denotes a single probability. An evidence \( e \) is an observed assignment of values for some or all of the random variables in the network. An
evidence $e$ for all the context nodes $C$ is an observation for a certain context; for example, Vehicles density has value low and Obstacle has value false. For simplicity, we use the term context also to refer to the associated evidence in the Bayesian Network.

In this paper, since we focus on revising the sanction that enforces a norm, we consider Norm Bayesian Networks in which the set of norm nodes $N$ is composed by a single node, like in Figure 3. This node represents the only norm that is enforced in the system. Since the norm is never disabled, in the following we ignore the disabled value of the corresponding node in the Bayesian Network.

### 3.2 Norms and Overall Objectives in MAS

Consider a set of agent types $T$, each type corresponding to a rational preference as per Section 2. Also, take a set of agents $A_g = \{a_1, \ldots , a_n\}$, each with a specific type from $T$. We use $Pref(a) \in T$ to indicate that agent $a \in A_g$ behaves according to a type from $T$. Moreover, we assume that the actions performed in a multiagent system are uniformly distributed over all the agents: no agent performs more actions than others. In the ring road scenario, every car performs an action—setting its speed—at every time instant. Since the cars in the ring road do not change, each car performs the same number of actions in any time window.

Given these assumptions and a norm $N$, we say that $N$ is well defined if the probability that $N$ is violated, denoted as $P(N_{\text{viol}})$, is never higher than the percentage of agents in the MAS with a reason to violate $N$. In other words, the upper bound of the probability $P(N_{\text{viol}})$ is the percentage of the agents with a reason to violate $N$.

Let $N$ be a norm, $T = \{t_1, \ldots , t_k\}$ be a set of agent types in the multiagent system, and $\delta = (d_1, \ldots , d_k)$ be a distribution over the agent types where $d_i$ is the percentage of population of agents of type $t_i$ and $\sum_{i=1}^{k} d_i = 1$. The percentage of agents with a reason to violate $N$ is $\sum_{i=1}^{k} (d_i \cdot \text{hasReason}(i, N))$, with hasReason($i, N$) = 1 if agent type $t_i$ has a reason to violate $N$, 0 otherwise. Note that, since we represent agent types as rational preferences, increasing the sanction $s$ of a norm $N = (p, s)$ does not increase the percentage of agents with a reason to violate $N$. Therefore, given $k$ agent types and $\max_B(T,N)$ as the maximum budget among all agent types to violate a well-defined norm $N = (p, s)$, the percentage of agents with a reason to violate a well-defined norm $N' = (p, \max_B(T,N))$ is 0. This is to say that increasing the sanction of a norm to the maximum budget that any agent is willing to pay cause all agents to comply with the norm.

Consequently, given two well-defined norms $N = (p, s_1)$ and $N' = (p, s_2)$ such that $s_2 > s_1$, the upper bound of the probability $P(N'_{\text{viol}})$ is never bigger than the upper bound of the probability $P(N_{\text{viol}})$. Consider as an example a norm $N$, a set of three agent types $T = \{t_1, t_2, t_3\}$ and a population of agents where 60% of them have type $t_1$, 30% has type $t_2$ and the remaining 10% has type $t_3$. Suppose that each agent $a$ with type $t_1$ has a maximum budget to violate $N$ equals to 2 (i.e., $\max_B(a, N) = 2$), agents with type $t_2$ have maximum budget 3 and agents with type $t_3$ have maximum budget 0. The maximum budget among all agents types for violating $N$ is 3, i.e., $\max_B(T,N) = 3$. Figure 4 reports the upper bound of the probability of violating $N$ for this example with different sanctions.

![Figure 3: A Norm Bayesian Network for the ring road.](image1.png)

![Figure 4: Example of upper bound of the probability violating a well-defined norm $N$ w.r.t. the sanction associated to $N$ and agents' preferences.](image2.png)

Note that the upper bound of $P(N_{\text{viol}})$ describes a hypothetical situation where all agents behave according to their preferences, no contextual factor influences agent behavior, and interactions among agents do not prevent them to act according to their preferences. This would happen, for example, when a single car drives on an empty highway with perfect road and car conditions. However, the actual probability to violate a norm is affected by the agent interactions and by the MAS environment. Even if all agents have a reason to violate a norm, due to their interaction or to environmental circumstances (e.g., large number of cars in the ring-road), none of them may end up violating it. We call the monitored probability of violating (obeying) a norm exhibited norm violation (obedience).

The concept of well-defined norm concerns the relationship between a norm and agent preferences. In a multiagent system, norms are enforced to achieve some system-level objectives. We define here two properties that relate a norm with the overall objectives: the concept of synergy between a norm and the overall objectives, and the concept of effectiveness of a norm.

We say that there is a positive synergy between a norm and the overall system-level objectives if it is more likely to achieve the overall objectives when the norm is obeyed than when it is violated. A positive synergy between a norm $N$ and a set of boolean objectives $O$ exists if $P(O_{\text{true}}|N_{\text{ob}}) > P(O_{\text{true}}|N_{\text{viol}})$. We say that there is a negative synergy between $N$ and $O$ if $P(O_{\text{true}}|N_{\text{ob}}) < P(O_{\text{true}}|N_{\text{viol}})$. Finally, we say that there is no synergy between $N$ and $O$ if $P(O_{\text{true}}|N_{\text{ob}}) = P(O_{\text{true}}|N_{\text{viol}})$.

We say, instead, that a norm $N$ is effective if, when $N$ is enforced, $N$ guarantees the desired achievement level $\tau$ of the system-level objectives, i.e., when $P(O_{\text{true}}) > \tau$, with $P(O_{\text{true}}) = P(O_{\text{true}}|N_{\text{ob}}) \cdot P(N_{\text{ob}}) + P(O_{\text{true}}|N_{\text{viol}}) \cdot P(N_{\text{viol}})$.
Notice that, although setting sanctions to $\max B(T, N)$ makes all the agents compliant (i.e., $P(N_{\text{viol}}) = 0$ and $P(N_{\text{obs}}) = 1$), it does not necessary guarantee the achievement of the overall system objectives, as norms can be ineffective when obeyed by all agents.

The exhibited norm obedience, the synergy and the effectiveness are hard to determine while designing a MAS, due to the complexity of the system, the interaction between autonomous agents, and the uncertainty of the environment. However, they can be learned at runtime by monitoring the system performance. We follow the approach from [10] to learn such properties by means of the Norm Bayesian Network. We combine these properties with the agents’ preferences to revise the sanction of an ineffective norm $N$.

4 NORM REVISION

We propose two heuristic strategies for the revision of the sanction of a well-defined norm whose enforcement is currently ineffective. The two strategies that we propose, called synergy and sensitivity, leverage the knowledge learned at runtime about norm effectiveness and the knowledge about the preferences of the agents. Below, our explanation assumes a specific context $c$.

Take the Norm Bayesian Network in Figure 3. By analyzing the conditional probability tables (CPTs) of the objectives nodes $O = \{\text{TripDur, Halted}\}$, we can determine whether norm $N$ is effective or not in context $c$. If $N$ is not effective $(P(O_{\text{true}}) < \tau)$, a revision of $N = (p, s)$ is triggered. Here, we aim to revise the sanction $s$. For example, if the norm (speed, 50) is ineffective with an obstacle and high vehicle density, we aim to identify another value for the sanction. $\max B(T, N)$ among all agent types in $T$ to violate a well-defined norm $N$, the set of possible sanctions that can be used to enforce $N$ is $S = \{s \in N | s \leq \max B(T, N) + 1\}$. This set is the search space within which the two strategies synergy and sensitivity search for a new sanction for an ineffective norm.

**Synergy:** Sanction revision based on norm-objectives synergy. This strategy uses the learned information about the synergy between $N$ and the objectives $O$ in $c$ and about the exhibited norm obedience. If there is a positive synergy between $N$ and $O$ in $c$, the objectives $O$ are more likely to be achieved when $N$ is obeyed. In this case, by reducing the violations of $N$ we should expect to increase $P(O_{\text{true}})$. In case there is a negative synergy between $N$ and $O$ in $c$, instead, we should expect that reducing the violations of $N$ would also increase $P(O_{\text{true}})$ by increasing the violations of $N$. We aim at achieving the required change by revising the sanction of norm $N$. In synergy, we define the new sanction $s'$ as the closest sanction to $s$ in the set $S$ that has never been attempted before and that is expected to increase (or decrease) $P(N_{\text{viol}})$. Note that, by choosing the closest sanction, synergy favors stability in the normative system of the MAS, i.e., it applies minimal changes to the norm sanction.

**Sensitivity:** Sanction revision based on sensitivity analysis. This strategy is based on the sensitivity analysis technique from probabilistic reasoning [6]. We aim not only at determining the direction of the revision—i.e., increasing or decreasing the probability $P(N_{\text{viol}})$ as in the case of synergy—but also at providing an estimation of the required change in the probability $P(N_{\text{viol}})$ in order to make $N$ effective in context $c$. The probability $P(N_{\text{viol}})$ is a parameter $\theta_{N_{\text{viol}}}$ of the Norm Bayesian Network. Given the node $N$, we want to identify possible changes to the parameter $\theta_{N_{\text{viol}}}$ that can ensure the satisfaction of the constraint $P(O_{\text{true}}) \geq \tau$.

We call required revision strength (RRS) the desired change $\Delta \theta_{N_{\text{viol}}}$ in the parameter $\theta_{N_{\text{viol}}}$ that ensures $P(O_{\text{true}}) \geq \tau$. Such a value can be determined by performing sensitivity analysis for the Bayesian Network [6], thus solving the following inequality:

$$P(O_{\text{true}}) + \frac{\delta P(O_{\text{true}})}{\delta \theta_{N_{\text{viol}}}} \cdot \Delta \theta_{N_{\text{viol}}} \geq \tau \quad (1)$$

Considering the topology of the Norm Bayesian Network of Figure 3, with a single norm node (we will focus on the case of multiple norms in future work), we follow Chan et al. [6] and compute the derivative $\frac{\delta P(O_{\text{true}})}{\delta \theta_{N_{\text{viol}}}}$ as follows.

$$\frac{\delta P(O_{\text{true}})}{\delta \theta_{N_{\text{viol}}}} = \frac{P(O_{\text{true}}|N_{\text{viol}})}{P(N_{\text{viol}})} - \frac{P(O_{\text{true}}|N_{\text{obs}}, c)}{P(N_{\text{viol}})} \quad (2)$$

For the Bayesian Network of Figure 3, such derivative can be further simplified as $P(T\text{ripDur}_{\text{true}}|N_{\text{viol}}, c) \cdot P(Halted_{\text{true}}|N_{\text{viol}}, c) - P(T\text{ripDur}_{\text{true}}|N_{\text{obs}}, c) \cdot P(Halted_{\text{true}}|N_{\text{obs}}, c)$.

The calculated RRS determines the change in $P(N_{\text{viol}})$ that is estimated, based on data acquired from execution, to be required in order to make $N$ effective. Given a norm $N = (p, s)$ and the required revision strength RRS, we need to find a new norm $N' = (p', s')$ such that the applied revision strength (i.e., the difference $UB(N'_{\text{viol}}) - P(N_{\text{viol}})$ between the estimated upper bound $UB$ for the norm violation with $N'$ and the current exhibited norm violation with $N$) is as close as possible to RRS and such that $s'$ is as close as possible to $s$.

After enforcing the new norm $N'$ obtained by selecting a new sanction, we monitor the new behavior of the agents and detect the new exhibited norm violation $P(N'_{\text{viol}})$. We call actual revision strength the difference $P(N'_{\text{viol}}) - P(N_{\text{viol}})$ between the exhibited norm obedience with $N'$ and with $N$.

5 EXPERIMENTS

To evaluate our proposed norm revision strategies, we use a simulation of the ring road scenario of Section 2. We consider the two contextual variables Vehicle density, which can be low (40 cars in the ring road) or high (100 cars); and Obstacle, which is true when a halted car is placed on the outer lane of the ring road. Each car in the simulation is an agent that acts according to its preference. At each simulation step, every agent determines its desired speed in the context of the norm currently enforced, as described in Section 2. We consider the set of propositional atoms $L = \{15, 8, 3\}$ (where each element in $L$ stands for speed, i m/s) and the set of sanctions $S = \{0, 1, 2, 3\}$ for defining norms and agent preferences. We consider four types of agents:

- **BraveRich** prefers to drive fast and pay a maximum budget of 2, i.e., $(15, 0) \geq (15, 1) \geq (15, 2) \geq (8, 0) \geq (8, 1) \geq (8, 2) \geq (3, 0) \geq (3, 1) \geq (3, 2)$;
- **BraveMiddleClass** prefers to drive fast but pay a maximum budget of 1, i.e., $(15, 0) \geq (15, 1) \geq (8, 0) \geq (8, 1) \geq (3, 0) \geq (3, 1) \geq (15, 2) \geq (8, 2) \geq (3, 2)$;
We monitored the behavior of the cars and sanctioned each car that violated the currently enforced norm. A sanctioned car was not sanctioned anymore until it completed a full round of the ring road.

We experiment with three distributions of agents: uniform (the entire population of agents is uniformly distributed across the four types), mostly compliant (75% of agents belong to type Cautious and the rest is uniformly distributed across the remaining types), and mostly violating (75% of agents belong to type BraveRich and the rest is uniformly distributed across the remaining types). We enforce two different norms: $N = (3, s)$ and $N = (8, s)$ with $s \in \mathcal{S}$ in all the four possible operating contexts, with all the three agents distributions. Figure 5 illustrates the upper bounds of the probability of violating the two norms above defined (as per Section 3.2) for the three agent type distributions.

Figure 5: Upper bound of the probability of violating norms $(3, s)$ and $(8, s)$ with different agent type distributions.

During the simulation, we collected data about norms obedience and objectives achievement in the four different operating contexts. We monitored the behavior of the cars and sanctioned each car that violated the currently enforced norm. A sanctioned car was not sanctioned anymore until it completed a full round of the ring road. The boolean value of the objectives was measured every 50 steps.

The theoretical trip time is $\tau(N) = \frac{d_i t_{i, N}}{t_{i, N}}$, where $d_i$ is the percentage of agents of type $t_i$, and $t_{i, N}$ is the theoretical time needed by $t_i$ to complete a round in case of free ring road when norm $N$ is enforced.

- BravePoor prefers to drive fast but does not want to pay any sanction, i.e., $(15, 0) \geq (8, 0) \geq (3, 0) \geq (15, 1) \geq (8, 1) \geq (3, 1) \geq (15, 2) \geq (8, 2) \geq (3, 2)$;
- Cautious prefers to drive slow and be compliant, i.e., $(3, 0) \geq (8, 0) \geq (15, 0) \geq (3, 1) \geq (8, 1) \geq (15, 1) \geq (3, 2) \geq (8, 2) \geq (15, 2)$.

The boolean value of the objectives was measured every 50 steps. The theoretical trip duration is achieved if, on average, less than 10% of cars were halted in the ring road.

To evaluate the theoretical average trip time as two possible scenarios (the 28.5% of the 256 configurations that are above defined in Section 3.2) for the three agent type distributions. The system has $4^3 = 256$ possible configurations: 4 possible sanctions for norm $N$ in any of the 4 contexts. We defined a different $\tau$ for each of the six scenarios—two norms enforced in three agent type distributions—based on an exploration of the distribution of the 256 configurations. Figure 6 shows this distribution and highlights how in each scenario the optima are differently located in the hill climbing search space. Thus, the six test scenarios are independent, thereby increasing the generality of our results.

5.1 Analysis of the Results

Table 1 reports results in terms of convergence speed for the six scenarios. Visual inspection highlights that our heuristics, and in particular sensitivity, outperform the baseline uninformed heuristics. Combining all scenarios, our informed heuristics required on average 9.85 steps to find an optimal configuration (sensitivity only 4.98) as opposed to 92.39 steps for the uninformed heuristics.

In the six tested scenarios, configurations that are close to each other in the search space do not necessarily have similar outcomes. This suggests that a good search strategy should weight more exploration than exploitation (i.e., diversifying the configurations to be tried should be weighted more than refining the search in the vicinity of the current configuration).

Among the uninformed heuristics, S10 appeared to be the best one. In particular, this strategy was successful in scenarios where the number of optimal solutions was higher. See, for example, scenario $Uniform$ and $N = (3, s)$, where there are 73 optimal configurations (the 28.5% of the 256 configurations that are above $\tau = 0.7$), or scenario $Mostly$ $violating$ and $N = (8, s)$, which contains 65 optimal configurations, 25% of all the 256 configurations. In these cases, the balance between exploration and exploitation employed by S10 allowed to find an optimal configuration by trying less than 5% of the possible configurations (see also Figure 7, which compares also the average percentage of explored configurations). Conversely, the heuristic D8, defining a broader neighborhood for the configurations and favouring exploitation over exploration, required significantly more steps than all the others in the six scenarios.

1For instance, the distance between the two configurations $\{(low\rightarrow true, 1), (low\rightarrow false, 2), (high\rightarrow true, 0), (high\rightarrow false, 3)\}$ and $\{(low\rightarrow true, 3), (low\rightarrow false, 2), (high\rightarrow true, 1), (high\rightarrow false, 0)\}$ is 6. The choice of value 8 allows revisions with a maximum average distance of 2 for each sanction.

2Notice that such percentage can be lower than the number of steps performed by hill climbing, for the number of steps also counts the backtracking operations.

3For instance, the percentage of agents of type $t_i$, and $t_{i, N}$ is the theoretical time needed by $t_i$ to complete a round in case of free ring road when norm $N$ is enforced.
Table 1: Comparison of the strategies in terms of average number of steps required to find an optimal solution.

<table>
<thead>
<tr>
<th>Heuristic</th>
<th>Average number of steps (σ)</th>
<th>uniform (N = 3, s)</th>
<th>mostly compliant (N = 3, s)</th>
<th>mostly violating (N = 3, s)</th>
<th>mostly compliant (N = 8, s)</th>
<th>mostly violating (N = 8, s)</th>
<th>Total average</th>
</tr>
</thead>
<tbody>
<tr>
<td>D8</td>
<td></td>
<td>174.45 (109.65)</td>
<td>230.51 (60.27)</td>
<td>236.44 (46.18)</td>
<td>199.70 (96.49)</td>
<td>210.52 (60.27)</td>
<td>209.44 (87.21)</td>
</tr>
<tr>
<td>S10</td>
<td></td>
<td>8.73 (3.65)</td>
<td>49.41 (36.37)</td>
<td>28.94 (18.27)</td>
<td>59.44 (59.88)</td>
<td>28.52 (19.32)</td>
<td>31.48 (35.79)</td>
</tr>
<tr>
<td>S20</td>
<td></td>
<td>15.02 (9.72)</td>
<td>50.01 (31.24)</td>
<td>41.14 (27.55)</td>
<td>52.48 (42.16)</td>
<td>34.79 (20.89)</td>
<td>36.25 (29.96)</td>
</tr>
<tr>
<td>SYNERGY</td>
<td></td>
<td>15.80 (25.91)</td>
<td>3.04 (1.68)</td>
<td>50.00 (25.74)</td>
<td>1.25 (0.83)</td>
<td>11.96 (4.19)</td>
<td>14.73 (22.40)</td>
</tr>
<tr>
<td>SENSITIVITY</td>
<td></td>
<td>1.67 (1.29)</td>
<td>2.25 (1.26)</td>
<td>14.13 (22.73)</td>
<td>1.81 (1.26)</td>
<td>3.97 (2.48)</td>
<td>6.02 (3.98)</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 6: Average probability of objectives achievement for the 256 tried configurations in the 6 different tested scenarios.

Figure 7: Average percentage of explored configurations before finding an optimal one.

Differently from the uninformed heuristics, SYNERGY and SENSITIVITY define the neighborhood of the configurations by leveraging runtime execution data. In SYNERGY, the neighborhood of a configuration consists of the closest configuration that is expected to decrease (or increase, if negative synergy between N and O) the $P(N_{eq})$. This strategy was successful, as expected, in scenarios where there is a clear correlation between the norm being violated and the achievement of the overall objectives. See for example the results of the three scenarios with norm $N = (8, s)$. Such scenarios are similar and differ mostly in the number of optimal configurations: 17 optimal configurations above $\tau = 0.8$ (6.6% of all) in scenario Uniform and $N = (8, s)$, 49 above $\tau = 0.85$ (19% of all) in scenario Mostly compliant and $N = (8, s)$, and 65 (25% of all) in scenario Mostly violating and $N = (8, s)$. In these scenarios the optimal configurations are the ones where action is set to 3 in context low-true (i.e., all agents’ behaviors are compliant with $N$). In scenario Uniform and $N = (8, s)$ all agents must be compliant also in context high-true, while in scenario Mostly compliant and $N = (8, s)$ in context high-true the exhibited norm violation has to be $< 0.17$. In these cases there is a clear correlation between the obedience of norm $N = (8, s)$ and the achievement of the overall objectives. The results of SYNERGY confirm that such correlation can be successfully learned and exploited at runtime to determine optimal sanctions. Even though on average better than uninformed ones, however, SYNERGY performed poorly in the scenarios where the overall objectives could be achieved only when more restrictive constraints about norm violation were satisfied. For instance in scenario Mostly compliant and $N = (3, s)$, the only 10 optimal configurations are the ones such that the exhibited norm violation is $0.09 < p < 0.17$ in context low-true, $\leq 0.25$ in context high-true, and $> 0$ in the other contexts.

Conversely from SYNERGY, SENSITIVITY appeared to be successful both in the cases where there is a clear norm-objectives synergy and in more complex cases, like Mostly compliant and $N = (3, s)$. 1615
Despite the restrictive constraints, sensitivity outperformed all other heuristics. The heuristic selects the new sanction to enforce, based on an estimation of the required change in the amount of exhibited norm violation. This estimation, allows sensitivity to perform on average better than synergy and the uninformed strategies and to quickly determine the optimal sanctions. On average, sensitivity required 4.98 steps to converge to an optimal configuration of sanctions for 4 different contexts. Remarkably, in 3 out of 6 tested scenario sensitivity required on average only around 2 steps, i.e., it required to explore only around 1% of the possible configurations in order to find one of the optimal ones.

Finally, our informed heuristics have a smaller standard deviation than the uninformed ones. This suggests they are robust, although more research is needed. Moreover, in all the tested scenarios, we only considered four sanctions, determined based on the agent preferences. This low number of sanctions penalizes the sensitivity heuristic, since it forces the selection of a new configuration that is only an approximation of the required revision strength determined by the sensitivity analysis. Nevertheless, the strategy appears to be the most successful in all the tested scenarios, confirming the hypothesis that estimating the degree of the required norm violation change through sensitivity analysis is an effective strategy.

6 RELATED WORK

Many approaches in the literature focused on the the design-time construction of robust normative MASs. For example, several researchers proposed techniques for proving the correctness of normative systems through the model checking of formulas that describe liveness or safety properties [2, 9, 12]. These works are very useful for the initial design of a MAS, but they are not sufficient to cope with the runtime unpredictability of the system that arises from the autonomy and heterogeneity of the participating agents.

Some frameworks have been proposed that formalize norm dynamics and allow assessing the impact of changed norms on the specification of a MAS, i.e., whether the designed MAS will be norm compliant. Knubbout et al. [14] propose a dynamic logic to formally characterize the dynamics of state-based and action-based norms. Both in Knubbout’s work [13, 14] and in Alechina et al.’s approach [2], norm change is restricted to norm addition. In this paper, instead, we explore how norms can be altered by establishing adequate sanctions that, using knowledge about agent preferences, maximizes the satisfaction of the system objectives.

Aucher et al. [4] introduce a dynamic context logic to describe the operations of contraction and expansion of theories that occur when removing or adding new norms. Governatori et al. [11] investigate from a legal point of view the application of theory revision to legal abrogations and annulments. We leave a study of the impact of norm revisions on an existing MAS legal system to future work.

Norm approximation [3] is a concept related to revision. The concept of approximation is defined with respect to a specific monitor: an approximated norm is synthesized from the original one to maximize the number of violations that can be detected by an imperfect monitor. Here, we assume perfect monitors and we investigate how to best enforce a set of norms through revisions. A relevant future direction is to integrate the two approaches so that our technique fits the (common) practical case of a MAS with imperfect monitors.

Chopra et al. [7] study how agent preferences—expressed in terms of goals—interact with norms—represented as commitments. In particular, they propose a framework for the agents to adapt their behavior. We take an orthogonal approach, for we study how to change the norms without altering the agent construction.

7 CONCLUSIONS

For a MAS to achieve its system-level objectives, the complexity and unpredictability of the agent interactions and of the environment must be taken into account. When engineering such systems, the available knowledge of these dynamics is only partial and incomplete. Therefore, MASs need to be regulated at runtime.

We propose a regulatory mechanism that relies on norms with sanction, an effective mechanism to influence agent behavior and regulate a MAS [5]. In our approach, we automatically revise the sanctions that are employed to enforce the norms. To do so, we first interpret—through a Bayesian Network—runtime execution data in terms of how well certain norms contribute to the achievement of the system-level objectives in different operating contexts. Then, we suggest a revision of the sanctions using two different heuristic strategies. An evaluation through a traffic regulation simulation shows that our heuristics outperform uninformed heuristics in terms of how fast they identify an optimal solution, especially the heuristic based on sensitivity analysis [6].

This work paves the way for numerous future directions. An in-depth evaluation of the scalability and computational complexity of the presented approach is necessary to assess its suitability for MASs with many norms and sanctions. Our simple—yet extensible—language for representing norms and agents’ preferences can be extended to consider complex norm types beyond atomic propositions. Our population of agents was defined according to specific types. Future work should study the effect of agents that deviate from the prototypical agent type. An obvious extension of our study consists of considering multiple norms, and the possible many-to-many synergies that may exist between norms and objectives. Finally, the revision strategies should be extended to revise both the norm proposition and the sanction, and to synthesize new norms.

REFERENCES