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:
Davide Dell'Anna, Mehdi Dastani, and Fabiano Dalpiaz. 2019. Runtime Revision of Norms and Sanctions based on Agent Preferences. In *Proc. of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), Montreal, Canada, May 13–17, 2019,* IFAAMAS, 9 pages.

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

Proc. of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), N. Agmon, M. E. Taylor, E. Elkind, M. Veloso (eds.), May 13–17, 2019, Montreal, Canada.

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.

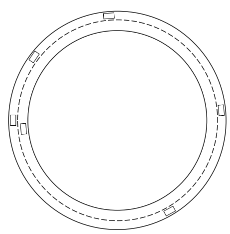


Figure 1: Two lanes ring road. Rectangles are vehicles, moving in counter clockwise direction.

The ring road is a simple example of a MAS. Although far from realistic traffic situations, the ring road illustrates the fundamental phenomena of emergent system-level properties, caused by the local decisions of individual agents, and the importance of mechanisms to control and steer such system-level behaviors.

We assume the main stakeholder of the ring road (the city council) has two system-level objectives: *to minimize the average time to complete a round of the ring road* and *to minimize the number of halted cars*. We consider two contextual variables that may influence the achievement of the system-level objectives, together with the vehicles' behavior: the *density of vehicles* and the *presence of an obstacle* in the ring road. To achieve the objectives, the behavior of the agents is regulated by enforcing norms concerning the speed limit, such as the norm *every vehicle in the ring road shall not exceed a speed of 50km/h, otherwise it will receive a sanction of 100€.* In other words, the ring road is a normative MAS.

Vehicles are autonomous agents, each belongs to a certain agent type that is characterized by the agent's preference. For instance a *cautious* agent is a type of agent that prefers to go slow rather than fast in the ring road. A *brave* agent is a type of agent that prefers to go fast rather than slow, even if it has to pay some money to do so.

2.2 Norms and Agents Preferences

In order to focus on the revision of the sanctions of the enforced norms, we propose a simple but extensible language for norms. Consider a set of propositional atoms $L = \{p_1, \ldots, p_k\}$, 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 $L = \{speed\_50, speed\_100\}$, a norm $N = (speed\_50, 100)$ 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 $Ag = \{a_1, \ldots, a_n\}$ be a set of agents. The preference of an agent a is denoted by $Pref(a) = (C, \geq)$, where $C = \{ (p_i, b_i) \mid 1 \leq i \leq k \;\&\; b_i \in \mathbb{N} \}$ and $\geq$ is a partial order on C. A pair $(p, b) \in Pref(a)$ 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 $(speed\_100, 0) \geq (speed\_100, 1) \geq (speed\_50, 0) \geq (speed\_50, 1)$ 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 if $(p_i = p_j)$ 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 $(speed\_50, 1) \geq (speed\_50, 0) \geq \ldots$, 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 $\geq = \geq_1 \cup \geq_2 \cup \{(c_1, c_2) \mid c_1 \in C_1 \;\&\; c_2 \in C_2\}$. 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. Without defining the notion of *violate* formally, we assume that a preferred state p_i (e.g., *speed*_100) violates a norm state p (e.g., *speed*_50) when p excludes p_i, and we write $viol(p_i, p)$ to indicate it. In the sequel, the pair (p_i, b) is said to be a violating pair w.r.t. a norm (p, s) when $viol(p_i, p)$; otherwise (p_i, b) is said to be a complying pair w.r.t. a norm (p, s).

Let $(C, \succeq)$ 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 $c_j \succeq c_i$ it holds that $viol(p_j, p)$ and $b < s$. For example, given the agent preference $(speed\_100, 0) \succeq (speed\_100, 1) \succeq (speed\_50, 0) \succeq (speed\_50, 1)$ and the norm $(speed\_50, 2)$, the most preferred pair to act upon is $(speed\_50, 0)$. Note that given a norm (p, s) it is possible to have the most preferred pair (p_i, b_i) such that $viol(p_i, p)$, 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 $(speed\_50, 1)$ that states "do not exceed 50km/h to avoid sanction 1", the most preferred pair to act upon is $(speed\_100, 1)$. 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 $Pref(a)$ is such that the most preferred pair is (p_i, b), and $viol(p_i, p)$.

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 $Pref(a) = (C, \succeq)$ 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 $maxB(a, N)$, 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: $\{ (p_i, b_i) \in C \mid (p_i, b_i) \succeq (p, b) \in C \}$, with (p, b) the first complying pair in $Pref(a)$. For example, consider $(speed\_100, 0) \succeq (speed\_100, 1) \succeq (speed\_50, 0) \succeq (speed\_50, 1)$ as the preference of agent a and $N = (speed\_50, 2)$ as the enforced norm. The maximum budget that a is willing to pay for violating N is *1*, i.e., $maxB(a, N) = 1$. Note that if the maximum buget 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 $(speed\_100, 2)$ 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.

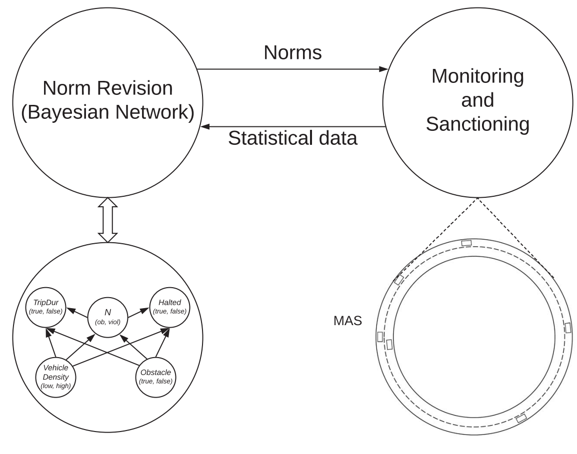


Figure 2: Illustration of the MAS supervision mechanism.

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* $\mathcal{NBN} = (\mathcal{X}, \mathcal{A}, \mathcal{P})$ [10] is a Bayesian Network where:

- $\mathcal{X} = \mathbf{N} \cup \mathbf{O} \cup \mathbf{C}$ are nodes that represent random variables in probability theory. $\mathbf{N}$, $\mathbf{O}$ and $\mathbf{C}$ are disjoint sets. $\mathbf{N}$ consists of *norm nodes*; each node $N \in \mathbf{N}$ cornensponds to a norm and has a discrete domain of 3 possible values: *obeyed*, *violated* and *disabled.* $\mathbf{O}$ consists of *objective nodes*; each node $O \in \mathbf{O}$ corresponds to a boolean objective and has a discrete domain of 2 values: *true* and *false.* Finally, $\mathbf{C}$ consists of *context nodes*; each node $C \in \mathbf{C}$ corresponds to a contextual variable and can have a discrete or continuous domain of values.
- $\mathcal{A} \subseteq (\mathbf{C} \times \mathbf{N}) \cup (\mathbf{C} \times \mathbf{O}) \cup (\mathbf{N} \times \mathbf{O})$ 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.
- $\mathcal{P}$ is a set of conditional probability distributions. These are encoded into conditional probability tables (CPTs), each one associated with a node in $\mathcal{X}$ 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 $(X, Y, \ldots)$ for random variables; bold uppercase $(\mathbf{X}, \mathbf{Y}, \ldots)$ 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; $\mathbf{O}_v$ denotes an assignment of value v to all nodes in $\mathbf{O}$; P denotes a single probability. An evidence $\mathbf{e}$ is an observed assignment of values for some or all of the random variables in the network. An evidence $\mathbf{c}$ for all the context nodes $\mathbf{C}$ 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 Network*s in which the set of norm nodes $\mathbf{N}$ 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.

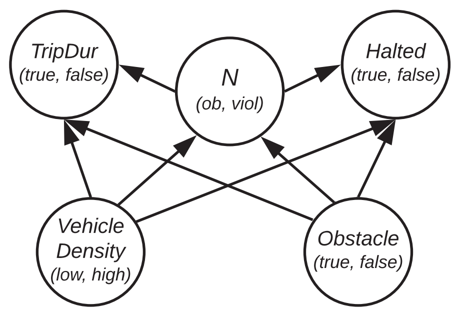


Figure 3: A *Norm Bayesian Network* for the ring road.

3.2 Norms and Overall Objectives in MAS

Consider a set of agent types $\mathcal{T}$, each type corresponding to a rational preference as per Section 2. Also, take a set of agents $Ag = \{a_1, \ldots, a_n\}$, each with a specific type from $\mathcal{T}$. We use $Pref(a) \in \mathcal{T}$ to indicate that agent $a \in Ag$ behaves according to a type from $\mathcal{T}$. 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 $P(N_{viol})$, 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 $P(N_{viol})$ is the percentage of the agents with a reason to violate N.

Let N be a norm, $\mathcal{T} = \{t_1, \ldots, t_k\}$ be a set of agent types in the multiagent system, and $\delta = (d_1, ..., d_k)$ be a distribution over the agent types where d_i is the percentage of population of agents of type t_i and $\forall i \in 1 \ldots k: \ d_i \in [0,1]$ and $\sum_{i=1}^{k} d_i = 1$. The percentage of agents with a reason to violate N is $\sum_{i=1}^{k}(d_i \cdot hasReason(i,N))$, 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 $maxB(\mathcal{T},N)$ 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 $N' = (p, maxB(\mathcal{T},N))$ 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 $P(N'_{viol})$ is never bigger than the upper bound of the probability $P(N_{viol})$. Consider as an example a norm N, a set of three agent types $\mathcal{T} = \{t_1,t_2,t_3\}$ and a population of agents where 60% of them has 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., $maxB(a,N)) = 2$), 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., $maxB(\mathcal{T},N) = 3$. Figure 4 reports the upper bound of the probability of violating N for this example with different sanctions.

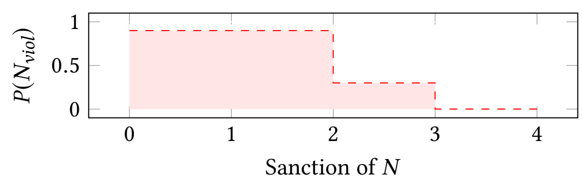


Figure 4: Example of upper bound of the probability of violating a well-defined norm N w.r.t. the sanction associated to N and agents' preferences.

Note that the upper bound of $P(N_{viol})$ 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 $\mathbf{O}$ exists if $P(\mathbf{O}_{true}|N_{ob}) > P(\mathbf{O}_{true}|N_{viol})$. We say that there is a *negative synergy* between N and $\mathbf{O}$ if $P(\mathbf{O}_{true}|N_{ob}) < P(\mathbf{O}_{true}|N_{viol})$. Finally, we say that there is *no synergy* between N and $\mathbf{O}$ if $P(\mathbf{O}_{true}|N_{ob}) = P(\mathbf{O}_{true}|N_{viol})$.

We say, instead, that a norm N is *effective* if, when N is enforced, N guarantees the desired achievement level τ of the system-level objectives, i.e., when $P(\mathbf{O}_{true}) > \tau$, with

$$P(\mathbf{O}_{true}) = P(\mathbf{O}_{true}|N_{ob}) \cdot P(N_{ob}) + P(\mathbf{O}_{true}|N_{viol}) \cdot P(N_{viol})$$

Notice that, although setting sanctions to $maxB(\mathcal{T}, N)$ makes all the agents compliant (i.e., $P(N_{viol}) = 0$ and $P(N_{ob}) = 1$), 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 $\mathbf{O} = \{TripDur, Halted\}$, we can determine whether norm N is effective or not in context **c**. If N is not effective ($P(\mathbf{O}_{true}|\mathbf{c}) < \tau$), a revision of $N = (p, s)$ is triggered. Here, we aim to revise the sanction s. For example, if the norm ($speed\_50$, 1) is ineffective with an obstacle and high vehicle density, we aim to identify another value for the sanction: 0, 2, 3, ... Given a norm N, a set of agent types $\mathcal{T}$ and the maximum budget $maxB(\mathcal{T}, N)$ among all agent types in $\mathcal{T}$ to violate a well-defined norm N, the set of possible sanctions that can be used to enforce N is $\mathcal{S} = \{s \in \mathbb{N} | s \leq maxB(\mathcal{T}, N) + 1\}$. 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 $P(\mathbf{O}_{true}|\mathbf{c})$. In case there is a negative synergy between N and **O** in **c**, instead, we should expect that increasing the violations of N would also increase of $P(\mathbf{O}_{true}|\mathbf{c})$ 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 $\mathcal{S}$ that has never been attempted before and that is expected to increase (or decrease) $P(N_{viol}|\mathbf{c})$. 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 $P(N_{viol}|\mathbf{c})$ as in the case of SYNERGY—, but also at providing an estimation of the required change in the probability $P(N_{viol}|\mathbf{c})$ in order to make N effective in context **c**. The probability $P(N_{viol}|\mathbf{c})$ is a parameter $\theta_{N_{viol|\mathbf{c}}}$ of the *Norm Bayesian Network*. Given the node N, we want to identify possible changes to the parameter $\theta_{N_{viol|\mathbf{c}}}$ that can ensure the satisfaction of the constraint $P(\mathbf{O}_{true}|\mathbf{c}) \geq \tau$.

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

$$P(\mathbf{O}_{true}|\mathbf{c}) + \frac{\delta P(\mathbf{O}_{true}|\mathbf{c})}{\delta\theta_{N_{viol|\mathbf{c}}}} \cdot \Delta\theta_{N_{viol|\mathbf{c}}} \geq \tau \tag{1}$$

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 $\frac{\delta P(\mathbf{O}_{true}|\mathbf{c})}{\delta\theta_{N_{viol|\mathbf{c}}}}$ as follows.

$$\frac{\delta P(\mathbf{O}_{true}|\mathbf{c})}{\delta\theta_{N_{viol|\mathbf{c}}}} = \frac{P(\mathbf{O}_{true}, N_{viol}|\mathbf{c})}{P(N_{viol}|\mathbf{c})} - P(\mathbf{O}_{true}|N_{ob}, \mathbf{c}) \tag{2}$$

For the Bayesian Network of Figure 3, such derivative can be further simplified as $P(TripDur_{true}|N_{viol}, \mathbf{c}) \cdot P(Halted_{true}|N_{viol}, \mathbf{c}) - P(TripDur_{true}|N_{ob}, \mathbf{c}) \cdot P(Halted_{true}|N_{ob}, \mathbf{c})$.

The calculated *RRS* determines the change in $P(N_{viol}|\mathbf{c})$ 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 $UB(N'_{viol}|\mathbf{c}) - P(N_{viol}|\mathbf{c})$ 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 $P(N'_{viol}|\mathbf{c})$. We call *actual revision strength* the difference $P(N'_{viol}|\mathbf{c}) - P(N_{viol}|\mathbf{c})$ 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 ring road according to its most preferred pair to act upon 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 i in L stands for $speed\_i$ m/s) and the set of sanctions $\mathcal{S} = \{0, 1, 2, 3\}$ 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., $(15, 0) \geq (15, 1) \geq (8, 0) \geq (8, 1) \geq (3, 0) \geq (3, 1) \geq (15, 2) \geq (8, 2) \geq (3, 2)$;

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

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 belongs to type *Cautious* and the rest is uniformly distributed across the remaining types), and *mostly violating* (75% of agents belongs 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.

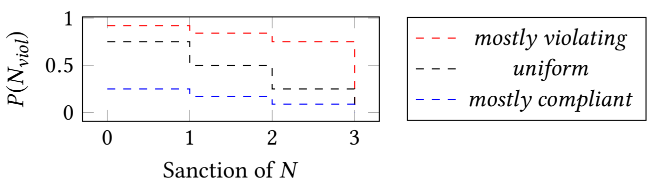


Figure 5: Upper bound of the probability of violating norms $(3, s)$ and $(8, s)$ with different agent type distributions.

During the simulation[1], 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 objective *Trip_Duration* was considered achieved if, on average in the 50 steps, the cars in the ring road took less than 2.5 times the theoretical average trip time[2] to complete a round of the ring road. The objective *Halted* was considered achieved if, on average, less than 10% of cars were halted in the ring road.

To evaluate SYNERGY and SENSITIVITY, we used the hill climbing implementation of the runtime norm supervision mechanism proposed in our previous work [10]. We evaluated the number of steps required by the hill climbing approach to converge to an optimal system configuration, i.e., an assignment of a sanction for the enforced norm to each of the system's operating contexts. An example of a configuration for the ring road is $\{(low\!-\!true, 1), (low\!-\!false, 2), (high - true, 0), (high - false, 3)\}$. We used our two strategies SYNERGY and SENSITIVITY as two possible *informed heuristics* for defining the neighborhood of a configuration, i.e., the configurations where the sanctions of the enforced norms are revised as suggested by the heuristics. We apply the hill climbing mechanism to compare the convergence speed of our informed heuristics against three *uninformed heuristics* that represented the baseline in [10]. Those heuristics do not take into account the runtime knowledge about norm effectiveness: *i.* maximum distance 8 (D8) includes in the neighborhood of a configuration all the configurations that are obtained by increasing or decreasing the sanctions of at most 8 sanction units[3]; *ii.* maximum size 10 (S10) and *iii.* maximum size 20 (S20) define a neighborhood that includes the 10 and 20 closest configurations to the current one, respectively. The used hill climbing approach stops when either all the possible configurations have been tried, or a configuration is found with an average objectives achievement probability above a threshold τ, so all the tested heuristics are always able to identify an optimal configuration.

To determine the average convergence speed to an optimal configuration—whose average objectives achievement probability is above τ—each strategy was executed starting from each possible configuration. The system has $4^4 = 256$ possible configurations: 4 possible sanctions for norm N in any of the 4 contexts. We defined a different τ 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 necessary 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[4]). 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 all the scenarios.

[1]For our experiments we used SUMO traffic simulator [15] and CrowdNav+RTX [16].
[2]The theoretical trip time is $\sum_{t_i \in \mathcal{T}} d_i \times t_{i,N}$, with $\mathcal{T}$ set of agent types, d_i the percentage of agents of type t_i, and $t_{i,N}$ the theoretical time needed by t_i to complete a round in case of free ring road when norm N is enforced.

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

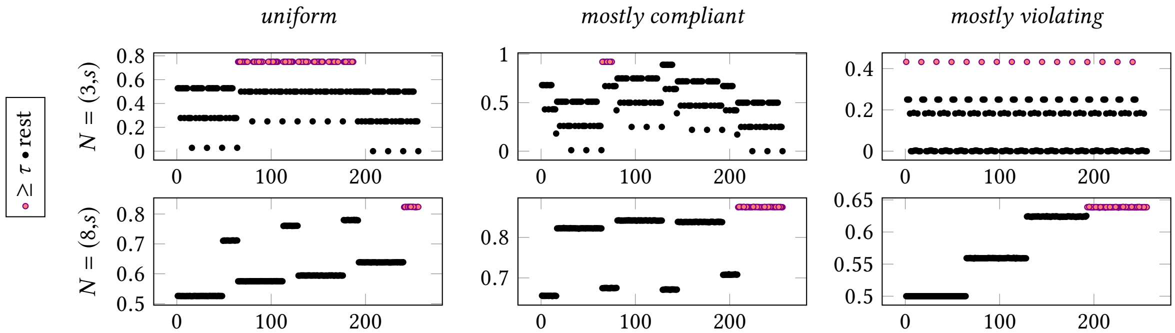
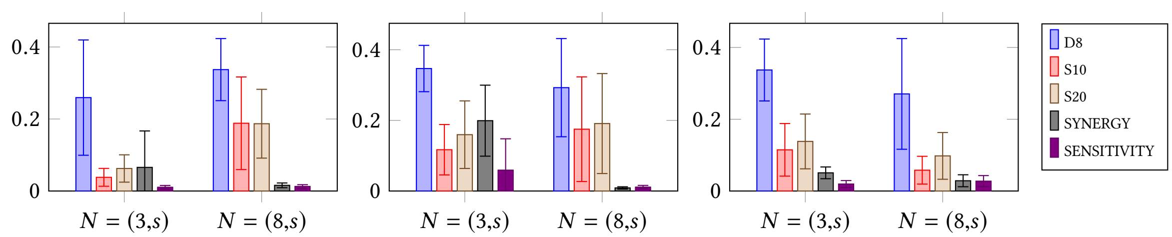


Figure 6: Average probability of objectives achievement for the 256 tried configurations in the 6 different tested scenarios.

Table 1: Comparison of the strategies in terms of average number of steps required to find an optimal solution.

	Average number of steps (σ)						
	uniform		*mostly compliant*		*mostly violating*		
Heuristic	$N = (3, s)$	$N = (8, s)$	$N = (3, s)$	$N = (8, s)$	$N = (3, s)$	$N = (8, s)$	Total average
D8	174.45 (109.65)	230.51 (60.27)	236.44 (46.18)	199.70 (96.49)	230.52 (60.27)	185.01 (107.29)	209.44 (87.21)
S10	8.73 (6.35)	49.41 (36.37)	28.94 (18.27)	59.44 (59.88)	28.52 (19.32)	13.84 (9.91)	31.48 (35.79)
S20	15.02 (9.72)	50.01 (31.24)	41.14 (27.55)	52.48 (42.16)	34.79 (20.89)	24.04 (16.67)	36.25 (29.96)
SYNERGY	15.80 (25.91)	3.04 (1.68)	50.00 (25.74)	**1.25 (0.83)**	11.96 (4.19)	6.32 (4.27)	14.73 (22.40)
SENSITIVITY	**1.67 (1.29)**	**2.25 (1.26)**	**14.13 (22.73)**	1.81 (1.26)	**3.97 (2.48)**	**6.02 (3.98)**	**4.98 (10.46)**



(a) ***uniform*** (b) ***mostly compliant*** (c) ***mostly violating***

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 netween N and **O**) the $P(N_{viol})$. 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 sanction 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 $0.09 \leq p \leq 0.17$ in context *low-true*, ≤ 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* $N = (3, s)$.

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. Knobbout *et al.* [14] propose a dynamic logic to formally characterize the dynamics of state-based and action-based norms. Both in Knobbout'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.

Cranefield *et al.* [8] present a Bayesian approach to norm identification. They show that agents can internalize norms that exist in an environment, by learning from the behavior that complies with or violates certain norms. This work constitutes a valuable addition to ours, for it shows that is is possible for agents to learn norms even when they are not explicitly communicated to them.

Tumer *et al.* [19] use multi-agent reinforcement learning in a smart traffic simulation to determine the behavior of the car agents that maximizes the utility of the city designer and of the individual agents. Their interesting work focuses on regimentation; instead, we focus on enforcement that does not violate agents' autonomy.

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

[1] Natasha Alechina, Nils Bulling, Mehdi Dastani, and Brian Logan. 2015. Practical Run-Time Norm Enforcement with Bounded Lookahead. In *Proc. of AAMAS*. 443–451. http://dl.acm.org/citation.cfm?id=2772937
[2] Natasha Alechina, Mehdi Dastani, and Brian Logan. 2013. Reasoning about Normative Update. In *Proc. of IJCAI*. 20–26. http://www.aaai.org/ocs/index.php/IJCAI/IJCAI13/paper/view/6884
[3] Natasha Alechina, Mehdi Dastani, and Brian Logan. 2014. Norm approximation for imperfect monitors. In *Proc. of AAMAS*. 117–124.

[4] Guillaume Aucher, Davide Grossi, Andreas Herzig, and Emiliano Lorini. 2009. Dynamic Context Logic. In *Proc. of LORI*. 15–26.

[5] Nils Bulling and Mehdi Dastani. 2016. Norm-based mechanism design. *Artificial Intelligence* 239 (2016), 97–142. https://doi.org/10.1016/j.artint.2016.07.001

[6] Hei Chan and Adnan Darwiche. 2004. Sensitivity Analysis in Bayesian Networks: From Single to Multiple Parameters. In *UAI '04, Proceedings of the 20th Conference in Uncertainty in Artificial Intelligence, Banff, Canada, July 7-11, 2004*. 67–75.

[7] Amit K. Chopra, Fabiano Dalpiaz, Paolo Giorgini, and John Mylopoulos. 2010. Reasoning About Agents and Protocols via Goals and Commitments. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems*. 457–464.

[8] Stephen Cranefield, Felipe Meneguzzi, Nir Oren, and Bastin Tony Roy Savarimuthu. 2016. A Bayesian Approach to Norm Identification. In *Proc. of ECAI*. 622–629. https://doi.org/10.3233/978-1-61499-672-9-622

[9] Mehdi Dastani, Davide Grossi, John-Jules Ch. Meyer, and Nick A. M. Tinnemeier. 2009. Normative Multi-Agent Programs and Their Logics. In *Normative Multi-Agent Systems*. http://drops.dagstuhl.de/opus/volltexte/2009/1907/

[10] Davide Dell'Anna, Mehdi Dastani, and Fabiano Dalpiaz. 2018. Runtime Norm Revision using Bayesian Networks. In *Proceedings of the 21st International Conference on Principles and Practice of Multi-Agent Systems (PRIMA2018)*.

[11] Guido Governatori and Antonino Rotolo. 2010. Changing legal systems: legal abrogations and annulments in Defeasible Logic. *Logic Journal of the IGPL* 18, 1 (2010), 157–194. https://doi.org/10.1093/jigpal/jzp075

[12] Max Knobbout and Mehdi Dastani. 2012. Reasoning under compliance assumptions in normative multiagent systems. In *Proc. of AAMAS*. 331–340.

[13] Max Knobbout, Mehdi Dastani, and John-Jules Ch. Meyer. 2014. Reasoning about Dynamic Normative Systems. In *Proc. of JELIA*. 628–636. https://doi.org/10.1007/978-3-319-11558-0_46

[14] Max Knobbout, Mehdi Dastani, and John-Jules Ch. Meyer. 2016. A Dynamic Logic of Norm Change. In *Proc. of ECAI*. 886–894. https://doi.org/10.3233/978-1-61499-672-9-886

[15] Daniel Krajzewicz, Jakob Erdmann, Michael Behrisch, and Laura Bieker. 2012. Recent Development and Applications of SUMO - Simulation of Urban MObility. *International Journal On Advances in Systems and Measurements* 5, 3&4 (December 2012), 128–138.

[16] Sanny Schmid, Ilias Gerostathopoulos, Christian Prehofer, and Tomás Bures. 2017. Self-Adaptation Based on Big Data Analytics: A Model Problem and Tool. In *Proc. of SEAMS*. 102–108. https://doi.org/10.1109/SEAMS.2017.20

[17] Yuki Sugiyama, Minoru Fukui, Macoto Kikuchi, Katsuya Hasebe, Akihiro Nakayama, Katsuhiro Nishinari, Shin-ichi Tadaki, and Satoshi Yukawa. 2008. Traffic jams without bottlenecksâĂŤexperimental evidence for the physical mechanism of the formation of a jam. *New journal of physics* 10, 3 (2008), 033001.

[18] Bas Testerink, Mehdi Dastani, and Nils Bulling. 2016. Distributed Controllers for Norm Enforcement. In *Proc. of ECAI*. 751–759. https://doi.org/10.3233/978-1-61499-672-9-751

[19] Kagan Tumer, Zachary T Welch, and Adrian Agogino. 2008. Aligning social welfare and agent preferences to alleviate traffic congestion. In *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems-Volume 2*. International Foundation for Autonomous Agents and Multiagent Systems, 655–662.

[20] Michael J. Wooldridge. 2009. *An Introduction to MultiAgent Systems (2. ed.)*. Wiley.