4

Self-Learning Governance of Competitive Multi-Agent Systems

Michael Pernpeintner

OrcidID ✉0000-0001-6939-1028
Institute for Enterprise Systems (InES),
University of Mannheim, Germany
pernpeintner@es.uni-mannheim.de

Abstract. Multi-Agent Systems (MAS) are widely used as a succinct model for distributed systems with (partly or fully) autonomous components. Whenever these components do not intrinsically cooperate, but pursue their individual goals in a purely selfish way (*Competitive MAS*), there is a natural challenge to prevent undesirable and destructive system behaviour and to achieve system-level objectives.

While agent autonomy is an essential characteristic of an MAS and can therefore not simply be replaced with full control or centralised management without losing its core functionality, it is still possible to achieve a certain level of control by applying a suitable governance approach.

I am proposing a new solution for this challenge. My approach adds to the usual agent/environment structure of an MAS a Governance component which can observe publicly available information about agents and environment, and, in turn, has the right to restrict the action spaces of agents and thus prevent certain environmental transitions.

As opposed to most existing methods, this approach does not rely on any assumptions about agent utilities, strategies or preferences. It therefore takes into consideration the fundamental fact that actions are not always directly linked to genuine agent preferences, but can also reflect anticipated competitor behaviour, be a concession to a superior adversary or simply be intended to mislead other agents.

The present paper motivates and describes the approach, defines the scope of the PhD project and shows its current status and challenges.

Keywords: Multi-Agent System, Competition, Governance, Restriction.

4.1 Introduction

4.1.1 Motivation

An essential feature of Multi-Agent Systems is the fact that agents depend on each other: The way the system behaves is not defined by the actions of one individual agent, but rather by the combination of all actions [18]. Therefore, a single agent can never be certain about the result of a chosen action. This mutual influence leads to strategic behaviour and sometimes even seemingly erratic actions—especially when an agent is human—, and at the same time decouples *intended* and *observed* system behaviour.

Example 1. Consider an MAS consisting of two agents X and Y, two environmental states A (initial state) and B, and two actions 0 and 1 for each agent, resulting in the joint action set $\{00, 01, 10, 11\}$ (the joint action 10 means that the first agent, X, chooses action 1, while the second agent, Y, takes action 0). The transition function of the MAS is shown in Figure 4.1. Imagine now an observer who sees the following sequence of actions and transitions:

$$A \xrightarrow{10} A \xrightarrow{01} A \xrightarrow{00} B$$

The observer, as is does not know the preferences of X and Y, cannot tell from the observed facts if X wanted to stay at state A and changed its action from 1 in the first step to 0 in the second step because it anticipated Y's second action, or if X observed the uselessness of its first action and then tried another strategy to reach state B (and failed again). This shows that intentions are not immediately linked to observable behaviour, and, in particular, no preference order over the environmental states can be concluded.

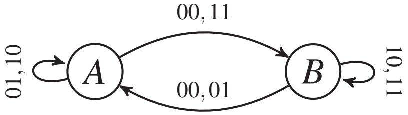

Fig. 4.1. Transition graph of a simple MAS

On the one hand, this is a challenge for a participating agent which needs to derive a strategy to counter its opponents' actions based on what it can see, but on the other hand, it makes it inherently hard to control or steer such a system using an external governing entity. I am specifically interested in the latter case, where there is a system-level objective (or "global desirable properties" [34]) to be achieved in addition to the individual goals of the agents.

It follows that preference elicitation (the process of deriving preferences over states from observed behaviour) is not feasible without additional assumptions about the link between actions and preferences. In general, the resulting preference order might be wrong, and relying on it could therefore lead to false conclusions about target conflicts and controlling decisions.

Nevertheless, the task of governing such an MAS requires some sort of planning and prediction of behaviour: In order to achieve a system-level objective, the Governance needs to prevent transitions which lead to violations of this objective. Therefore, it relies on collecting observable information and deriving knowledge about the future system behaviour. The two fundamental questions that it needs to answer based on this knowledge are: *"What will agents do next?"* and *"Which actions need to be forbidden in order to prevent undesirable transitions?"*

This online learning mechanism guarantees that the system can self-adapt to both changes in the setup–such as number of agents and system objective–and unforeseen strategic behaviour of the agents. It therefore ensures that the overall MAS is at the same time robust and flexible, without requiring manual intervention at run-time. From an Organic Computing perspective, moving the Governance logic and the selection of restrictions into the system makes the system "organic" in the sense that it can handle human agents in the same way as it handles agents based on experts systems, simple heuristics or sophisticated AI methods, and it therefore serves as a means for flexibly balancing the influences of otherwise uncontrolled agents.

4.1.2 Setting and Contribution

My PhD project is situated in the broader field of Organic Computing [27] and targets the problem of providing governance for competitive Multi-Agent Systems, purely based on the observation of public behaviour, i.e., actions and transitions. In contrast to most existing approaches for Governed MAS or Normative MAS, I argue that it is not reasonable to assume a-priori knowledge of agent utilities, preferences or strategies.

The contribution will consist of a new model for Governed MAS, proof of its feasibility and applicability, thorough analysis with respect to capabilities and complexity, and an evaluation which shows the performance of the framework in a real-world use case. Thereby, the research questions listed in Section 4.3.1 will be answered concisely and in depth.

4.1.3 Structure of the Paper

The remainder of this paper is organised as follows:

Section 4.2 defines the system model and the governing instance. Section 4.3 lists the research questions, shows what has already been accomplished and describes the necessary future work to complete the intended contribution. Section 4.4 recaps relevant existing work and places this project within the context of these approaches, while Section 4.5 outlines a real-world evaluation use case. Finally, Section 4.6 sums the paper up.

A more formal treatment of the multi-attribute case, including a governance algorithm and its evaluation, has recently been submitted [30]. Part of this submission is being included here in shortened form to show motivation, general system model, preliminary results and existing work.

4.2 Model

4.2.1 Agents and Environment

The general MAS model is based on [35]: Consider a finite set $\mathcal{P} = \{p_1, ..., p_n\}$ of agents (or players). An agent p_i perceives, at every time step $t \in \mathbb{N}_0$, the current state

$s_t \in \mathcal{S}$ of a temporally discretised environment and then acts within this environment by performing an action $a_i \in \mathcal{A}_i$, following a confidential (and not necessarily deterministic) *action policy* $\pi_i : \mathcal{S} \rightarrow \mathcal{A}_i$. The environmental state then changes from time step t to $t+1$ according to the combination of actions (the *joint action* $a = (a_1, ..., a_n) \in \mathcal{A}$) taken by the agents, as expressed by a *transition function* $\delta : \mathcal{S} \times \mathcal{A} \rightarrow \mathcal{S}$.

Definition 1. *A* Multi-Agent System *is the 6-tuple*

$$\mathcal{M} = (\mathcal{P}, \mathcal{S}, \mathcal{A}, \pi, \delta, s_0) \ .$$

4.2.2 Governance

In the basic MAS model of Section 4.2.1, the evolution of an MAS from t to $t+1$ follows the formula

$$s_{t+1} = \delta\left(s_t, \pi(s_t)\right) \ .$$

Since the action policies π_i are at the agents' sole discretion, one can see immediately that this progression can be influenced by an external authority via two levers only: Either by changing *what agents can do* (altering their action sets) or by changing *what consequences actions have* (altering the transition function).

The proposed governance model of this paper follows a strict separation of concerns: The transition function represents the unalterable evolution of the environment according to the actions taken by all agents, while the restriction of actions is performed by the Governance and therefore artificial. To use an analogy, the transition function accounts for the laws of nature in the system, whereas the Governance plays the role of the legislature.

4.2.2.1 Observation and Intervention

At the beginning of each cycle t, the Governance defines *allowed actions* before the agents choose their respective actions from this restricted action set:

$$\mathcal{A}_t = \Gamma_{s_{\mathcal{G}}^{(t)}}(s_t) \ ,$$

where $\mathcal{A}_t \subseteq \overline{\mathcal{A}}$ is a "rectangular" subset of a *fundamental action set* $\overline{\mathcal{A}} = \prod_i \overline{\mathcal{A}}_i$, i.e., $\mathcal{A} = \prod_i \mathcal{A}_i^{(t)}$ with $\mathcal{A}_I^{(t)} \subseteq \overline{\mathcal{A}}_i \ \forall i$. The subscript in $\Gamma_{s_{\mathcal{G}}^{(t)}}(s_t)$ hints to the fact that Γ implicitly uses as an input not only the current environmental state s_t, but also the internal state $s_{\mathcal{G}}^{(t)} \in \mathcal{S}_{\mathcal{G}}$ of the Governance, which includes the knowledge acquired so far. Since this is always the case, the subscript will henceforth be omitted for brevity. The shape of $\mathcal{A}_t$ needs to be rectangular for the simple reason that agents act independently in each step, which means that it is not possible to make conditional restrictions such as $\mathcal{A}_t = \{(a,x),(b,x),(a,y)\}$ since the Governance cannot, in this example, force p_1 to choose action a whenever p_2 chooses action y.

For each agent p_i, there is a *neutral action* $\varnothing_i \in \overline{\mathcal{A}}_i$ which cannot be deleted from the set of allowed actions. The resulting joint action $\varnothing$ is therefore always allowed.

As soon as all agents have made and communicated their choice of action $a = (a_i)_i \in \mathcal{A}_t$, the Governance can use the information gathered by observing the actions and the subsequent transition to learn about the agents and the effectiveness of Γ. This *learning step* is expressed as an update of the Governance's internal state which, in turn, will be used by Γ in the next step, i.e.,

$$s_{\mathcal{G}}^{(t+1)} = \lambda\left(s_{\mathcal{G}}^{(t)}, s^{(t)}, a\right) .$$

As opposed to some authors [4], I make no distinction between legal and physical power: An agent can choose only from the set of currently allowed actions (which might change from one step to the next), and it is not possible to disobey this rule. Nevertheless, the neutral action ensures that the system can operate with missing or invalid input coming from the agents—it simply uses $\varnothing_i$ as a fallback.

4.2.2.2 System Objective

As mentioned in Section 4.1.1, I assume that there is a certain system objective which is to be fulfilled, in addition to the agent-specific goals (and maybe conflicting with those agent goals). This way, the restriction mechanism of the Governance has the clear purpose of fulfilling this objective. Since the Governance has only probabilistic information about the agents' future actions, its objective needs to be compatible with probabilistic reasoning and therefore quantifiable.

While the system objective can be an arbitrary function from $\mathcal{S}$ to $\mathbb{R}$, there are two common types: Either minimising (or maximising) a numerical parameter, which can directly be expressed by $c_{\mathcal{G}}$, or dividing the state space into obeying states $\mathcal{S}_+$ and violating states $\mathcal{S}_- := \mathcal{S} \setminus \mathcal{S}_+$. In the latter case, the function

$$c_{\mathcal{G}}(s) := \mathbb{1}_{\mathcal{S}_-}(s) \tag{4.1}$$

describes a system objective which prefers all obeying states to all violating states by minimising $c_{\mathcal{G}}$. Therefore, the Governance will pursue an obeying state with minimal restriction of the agents.

Definition 2. *The* system objective *of an MAS $\mathcal{M}$ is defined as a* cost function $c_{\mathcal{G}} : \mathcal{S} \to \mathbb{R}$ *such that the Governance tries to reach and maintain a state of minimal cost.*

This cost function simply defines the preference of the Governance over the states of the environment; it does not necessarily correspond to a "real" cost.

The definition of a *Governed Multi-Agent System* is now that of an MAS, together with a specification of the Governance's behavior:

Definition 3. *A* Governed Multi-Agent System *(GMAS) is the 10-tuple*

$$\mathcal{M}_{\mathcal{G}} = \left(\mathcal{P}, \mathcal{S}, \overline{\mathcal{A}}, \pi, \delta, s_0, s_{\mathcal{G}}^{(0)}, c_{\mathcal{G}}, \Gamma, \lambda\right)$$

with $\Gamma : \mathcal{S} \to 2^{\overline{\mathcal{A}}}$ *and* $\lambda : \mathcal{S}_{\mathcal{G}} \times \mathcal{S} \times \mathcal{A} \to \mathcal{S}_{\mathcal{G}}$.

4.2.3 Run-time Process

The sequence of actions taken by the different components in one time step is shown in Figure 4.2. At each step, the Governance can define allowed actions via Γ (before the agents act) and learn from the observed actions via λ (after the agents have acted). The environment itself is not affected at all by the existence of the Governance.

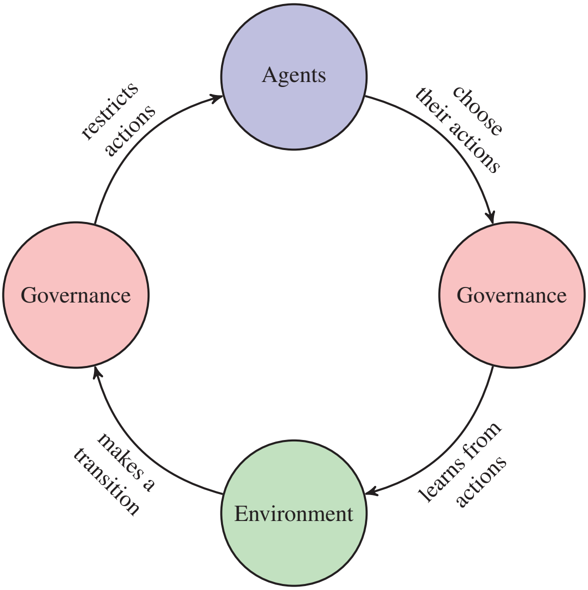

Fig. 4.2. Run-time Process

The performance of the Governance can now be measured by looking at two key parameters: (a) How high is the cost incurred at each state? and (b) How many restrictions were applied to achieve this cost? The second question naturally gives rise to the following notion:

Definition 4. *The* degree of restriction *of $\mathcal{G}$ at time t is the ratio of forbidden actions and fundamental actions:*

$$\mathfrak{r}_{\mathcal{G}}(t) := 1 - \frac{|\mathcal{A}_t|}{|\overline{\mathcal{A}}|} \in [0,1]$$

Taking this value as an indicator for Governance performance implies that all actions are equally important. Since this is not always the case, a more elaborate measure (e.g. comparing the size—with respect to some environment-specific metric—of the state set following from taking all actions in $\mathcal{A}_t$ and $\overline{\mathcal{A}}$) might be useful to better capture the “real” magnitude of the Governance-induced restrictions. This is a topic to be examined in future work.

Example 2. Consider a smart home environment consisting of 7 binary variables: $\mathcal{S} = T \times O \times W \times B \times H \times L \times A \cong \mathbb{B}^7$, where the variables denote Time (day/night), Occupancy (occupied/empty), Window (open/closed), Blinds (open/closed), Heating (on/off), Lights (on/off) and Alarm (on/off), respectively. n agents, who each have their individual preferences over the state, can now choose to change at most one of the variables W, B, H, L or A (the corresponding actions start at 1) at each step (they cannot, however, influence the Time or the Occupancy of the house). A variable is changed regardless of how many agents have chosen to change it at a single time step.

An exemplary progression of this system could be

$$\begin{aligned} s_0 = 1100101 &\xrightarrow{37\varnothing} 1110100 \xrightarrow{464} 1111110 \\ &\xrightarrow{\varnothing\varnothing 5} 1111010 \xrightarrow{564} 1110100 \xrightarrow{436} 1101110\,, \end{aligned}$$

where states are written as binary numbers and there are three agents acting upon the environment with the action sets

$$\overline{\mathcal{A}}_i = \{\varnothing, 3, 4, 5, 6, 7\}\ \forall i\,.$$

Time and Occupancy would of course need to be controlled by non-controllable environmental forces, but this is omitted here for simplicity.

Define now a Governance with cost function $c_{\mathcal{G}}$ as in Definition 2 where

$$\mathcal{S}_+ = \left\{ s \in \mathcal{S} : \left(\bar{w}(s) \vee \bar{h}(s)\right) \wedge \left(a(s) \vee o(s)\right) \wedge \left(\bar{l}(s) \vee o(s)\right) \right\}\,,$$

meaning that the system wants to make sure that (a) the window is not open while the heating is turned on, (b) the alarm is on when the house is empty, and (c) the lights are off when there's nobody home. It is therefore the task of the Governance to impose minimal restrictions on the agents while keeping $s_t \in \mathcal{S}_+$.

One can now see that $s_1 = 1110100$ incurs cost $c_{\mathcal{G}}(s_1) = 1$ since $s_1 \notin \mathcal{S}_+$. While the Governance probably cannot anticipate and prevent this transition between $t = 0$ and $t = 1$ due to lack of knowledge, it might be able to do so at a later time when enough information has been gathered. For example, at $t = 3$, the Governance could forbid action $5 \in \mathcal{A}_1$ such that the joint action 564 cannot happen. If p_1 now chooses action 3 instead, $s_4 = \delta(s_3, 364) = 1100000 \in \mathcal{S}_+$, and the Governance has therefore successfully prevented an undesirable transition.

4.3 Scope

4.3.1 Research Questions

The goal of this PhD project is the theoretical foundation, development, analysis and application of a GMAS platform which can be used to govern real-world Multi-Agent Systems with arbitrary agents. Therefore, the following research questions describe the gaps and open challenges in the current state of the art:

RQ1 Is the observation of actions and transitions, together with hard restriction of action spaces, sufficient and suitable for effective governance with respect to a given system objective? If not, which further assumptions, limitations or relaxations are necessary?

RQ2 Which data structures and algorithms can be used to create a scalable computation framework which can be used for online (real-time) governance? How does this framework perform in both benchmark and real-world applications?

RQ3 How can an agent (or a group of agents) manipulate the mechanism, and how can the Governance effectively identify and prevent manipulation?

4.3.2 Current Status

A widely accepted environmental limitation in MAS research is to assume a multivariate binary environment, i.e., $\mathcal{S} = \prod_{j=1}^{m} S_j$ for fixed $m \in \mathbb{N}$ and $S_j = \{s_j, \bar{s}_j\}$, such that $\mathcal{S} \cong \mathbb{B}^m$. This has the advantage of a compact representation; states can be written as Boolean arrays or encoded as natural numbers. I adopt this restriction for now, but keep in mind that my governance approach should, if possible, not be limited to this setting, but apply to (at least) arbitrary finite domains. I expect the permission of infinite or continuous domains or even irregular environmental "shapes" to pose new challenges, and will comment on this problem in Section 4.3.4.

Regarding actions and transitions, first assume that $\mathcal{A}_i \subseteq \{\varnothing, 1, \ldots, m\}$ and

$$\delta(s,a) = s' \text{ where } s'_j = \begin{cases} \bar{s}_j & \text{if } \exists i : a_i = j \\ s_j & \text{else} \end{cases}$$

which means that agents can choose to change one attribute per time step (or to do nothing, by choosing the neutral action $\varnothing$), and each attribute is toggled if at least one agent chooses to change it. As above, allowing more general environmental structures and more complex actions and transitions would cause additional challenges and require some additional assumptions. For example, aggregating agent actions with respect to a non-binary attribute [23] can be complex in itself: Does an agent request a certain value for a numerical attribute, or does it request a certain offset? Is the new value simply the mean of all requested values? Does it maybe only change when the agents can agree on a new value?

Theorem 1. *Let $\mathcal{M}$ be a GMAS with n agents, m binary attributes and q fundamental actions per agent. Then, for a given cost threshold $\alpha \geq c_{\mathcal{G}}(\delta(s_t, \varnothing)$, a Pareto-minimal restriction $\mathcal{A}_t \subseteq \overline{\mathcal{A}}$ can be computed in time $O\left(n^2 \cdot q^{(n+2)}\right)$.*

Proof. See [30].

Note that the complexity of this algorithm does not depend on the size of the environment, as long as the past observed actions per state are readily available. Therefore, it is suitable for MAS with large state spaces, but few actions—a typical scenario would be a video game where each player can take a constant (low) number of actions.

As shown in a first prototypical setup, it turns out that the smart home case (Example 2) can indeed be successfully governed by an algorithm built from Theorem 1. Figure 4.3 shows a part of the evaluation of [30] using a variable number of agents (2 = dotted line, 3 = dashed, 5 = continuous) which were set up with random state-action mappings and acted in this system for $0 \leq t \leq 100$. The chart shows a comparison between ungoverned and governed simulations, including the average cost for both simulations and the degree of restriction in the governed simulation. To minimise outliers, each line is the mean of 10 independent runs of the same simulation.

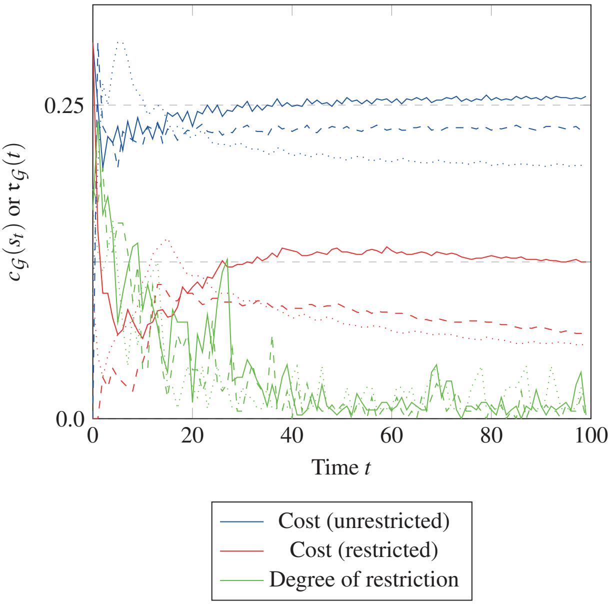

Fig. 4.3. Simulation of the Smart Home Example

4.3.3 Implementation

A meaningful evaluation of the theoretical approach is an essential ingredient for a PhD thesis which claims to provide a practical solution for governing competitive MAS. Great attention must thus be paid to an evaluation framework which allows for the testing of the approach as well as for a detailed comparison with competing approaches. Although the Governance component is the core of research and development, the overall performance and reliability also widely depend on realistic agents. If those agents are not immediately controlled by human players, there is still a need for strategic and "intelligent" behaviour in order to validate (or invalidate) the capabilities of the Governance.

In order to provide an optimal environment for development and testing, I am developing a Python-based Multi-Agent framework, specifically designed to be fed with different agent and governance functionalities. This framework provides measuring, logging and analysis of performance as well as direct comparison of different governance approaches (including no governance at all).

4.3.4 Challenges, Refinements and Extensions

While the general setting is very broad and applies to a wide range of MAS, I have made some restrictions and neglected particular issues so far in order to reduce complexity. Some of the topics which haven't been considered but are crucial for a deep understanding of Governed MAS are listed and explained in this section.

4.3.4.1 Fairness

In the current implementation (see Section 4.3.3), the Governance can effectively reduce its cost by defining restrictions based on an expected cost matrix. This approach forbids actions according to their expected cost impact, without taking into account previous restrictions or balancing the degree of freedom between agents. In extreme cases, the strategy can lead to some agents always being restricted to just one action, while others are not affected at all.

A natural question regarding this issue is whether "fairness" should be part of the Governance's decision process or even part of the system-level objective. If so, the concept of fairness needs to be well-defined in the context of MAS, and the Governance must be given a means to distinguish restrictions with respect to their evenness.

4.3.4.2 Derivation of Rules

The restriction function Γ is not required to provide any consistency, i.e., there is no link between $\mathcal{A}_t$ and $\mathcal{A}_{t+1}$ apart from the fact that both are subsets of $\overline{\mathcal{A}}$. Consequently, agents cannot anticipate what restrictions will be posed on their action space in the future. At the same time, the Governance does not justify its decisions or provide any reasons for them, but merely states what it allowed at the current step.

It might be useful to derive explicit rules or criteria for restricted and allowed actions, which could be expressed in a formal language. This would allow for better analysis of a system, for example regarding the link between agent behaviour and rule emergence. The field of Explainable AI deals with a similar issue of deriving abstract knowledge from sub-symbolic data.

4.3.4.3 Open agent sets

A typical problem with MAS is that agents, as they are autonomous entities, cannot be forced to do something. This implies that an agent might not react at all when it is asked to choose an action, or it might respond with incomprehensible or illegal data. In Section 4.2.2, a neutral action was introduced to cater for this fact—simply assume this action to be substituted for any invalid agent response. Nevertheless, the problem of agents spontaneously entering or leaving the system raises another question: Should the Governance treat all agents independently and individually? It might be a good idea to have a model which can handle unknown agents and apply some "general knowledge" to them, instead of assuming an empty knowledge base for each new agent. Such an approach would on the one hand free the Governance from having to identify and track each agent separately, and on the other hand allow it to (partly) carry over its knowledge to new agents joining the system.

4.3.4.4 Dynamic Agent Goals

It cannot, in general, be expected that agents remain consistent in their goals over the run-time of the system. In contrast, it is reasonable to assume that goals and strategies change gradually (not abruptly) over time. Therefore, the Governance should incorporate a mechanism which can deal with changing goals, for example by discounting old observations, or by categorising former observations according to consistency with the latest observed actions. This line of reasoning is closely connected to the field of belief revision [13].

4.3.4.5 Structure of environments and actions

When the environment consists of binary attributes and actions are merely toggling single attribute values, an MAS is fairly well-arranged. This, however, does not always represent the reality: There can be continuous or entangled environmental states, complex actions, non-trivial aggregation rules for different actions, and other complications. While the concrete implementation of a governance algorithm most likely depends on the choice of such properties, its general applicability should range over as large a class of systems as possible, and thus be able to deal with the general model from Section 4.2 instead of just binary multi-attribute MAS.

4.3.4.6 Distributed Governance

Multi-Agent Systems are one of the most common form of distributed systems, in which the overall computation task is carried out by independent entities which do not require central control and not even global information. Since this is a major asset of such systems, it seems counterproductive to add a central Governance which needs to aggregate and evaluate all agent actions at every step in order to do its job.

As a consequence, I will look at parallelising the Governance in order to ensure scalability. It seems that much of its work can be executed in a map-and-reduce fashion, but the existing algorithms haven't been designed according to this paradigm yet.

4.4 Related Work

The bulk of Multi-Agent research deals with the task of teaching agents how to act [17, 31], both in the cooperative case where there is a common goal and in the competitive case where conflicts are inherent. In contrast, I take the viewpoint of an outside entity wanting to "guarantee the successful coexistence of multiple programs" [37], that is, to define a degree of success and then influence it via suitable actions. Multi-Agent Systems can be classified with respect to this criterion as shown in Figure 4.4:

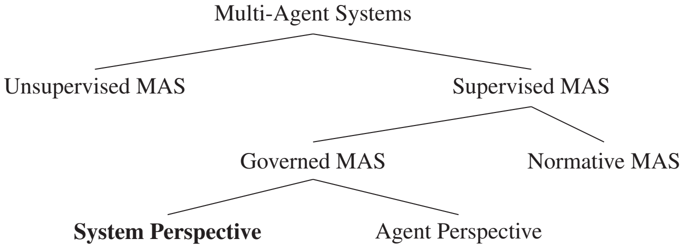

Fig. 4.4. Classification of Multi-Agent Systems

An MAS can either have a supervising entity which interferes with the agents in order to achieve a system objective, or this goal is achieved solely by the interaction of the agents (self-organisation and/or emergence [41], [26]).

When there is a supervisor, its decisions can be either binding (which I will call a *governed* MAS) or non-binding (normative). I follow here the reasoning of [4] who state that norms are "a concept of social reality [which does] not physically constrain the relations between individuals. Therefore it is possible to violate them." Note that this terminology is far from being unambiguous; for instance, [29] use the term *Normative Synthesis* for the *enforcement* of equilibria.

There are two perspectives of a Governed MAS: The viewpoint of a participating agent and that of the governing instance. In the latter case, the key points of interest are the level of control (or level of satisfaction of the system objectives) that can be achieved, and the necessary intervention.

There are many approaches developed from an agent perspective which can partly be applied to the system point of view, e.g., opponent modelling and Multi-Agent reinforcement learning. However, only few areas (e.g. *Normative Multi-Agent Systems* [5]) have been thoroughly examined from an observer's angle.

[17] and [16] identify two main research streams for competitive Multi-Agent Learning: Game theoretic approaches including auctions and negotiations, and Multi-

Agent Reinforcement Learning [36]. The latter add a layer of complexity to classical reinforcement learning [9], since competitive agents all evolve at the same time and therefore disturb the learning process of their opponents (*moving-target problem*) [28]. Both surveys, however, restrict their scope to learning agents, instead of external entities learning *about* agents.

Game theory in this context oftentimes deals with small, well-defined (and mostly contrived) scenarios [3, 15, 38] like two-player games with a fixed payoff matrix, which can be formally examined and sometimes even completely solved in terms of optimal responses and behavioural equilibria. What these solutions lack is widespread applicability to real-world settings where information is incomplete, environments are large and agents do not behave predictably. Therefore, the gap between academic use cases on the one hand and industrial and societal applications on the other hand is still large.

[37] realised that social laws can be used by designers of Multi-Agent Systems to make agents cooperate without controlling the agents themselves. They describe an approach to define such laws off-line and keep them fixed for the entire run-time of the system, and they mention the possibility that their laws are not always obeyed by the agents. From this reasoning, the two notions of *hard norms* and *soft norms* [33, 34] have emerged—the two categories which I call Governed MAS (GMAS) and Normative MAS (NMAS), respectively [19].

[34] argue that "achieving compliance by design can be very hard" due to various reasons (e.g. norm consistency and complexity of enforcement). Therefore, they reach the conclusion that NMAS are more suitable for open and distributed environments. In turn, the lack of hard obligations leads to concepts like sanctions, norm revision, norm conflict resolution, and others. NMAS have been researched from various perspectives and with various theoretical frameworks, among them formal languages and logics [7, 12, 29], Bayesian networks for the analysis of effectiveness [11], bottom-up norm emergence [26], and online norm synthesis [25]. Many of these approaches are also partially applicable to Governed MAS, but require adaptation and generalisation.

Another well-known problem of MAS is scalability [17, 40], especially for large state spaces. While the number of states is obviously exponential in the number of environmental variables, reasonable additional assumptions about the dependencies between variables can lead to much more compact representations of knowledge regarding preferences and utilities. Famously, this reasoning has been applied in the development from Q-learning [39] to Deep Q-learning [24]. While Q-tables and the corresponding Neural Networks describe the expected payoff of an action at a given environmental state (from an agent perspective) and hence define the choice of the next action, I need to describe the probability distribution of an action set, given an environmental state (from an observer's perspective).

Regarding preference orders over a set of alternatives, CP-nets [6] are among the most common data structures for encoding partial orders and enriching given knowledge with observations. They have been used extensively for preference aggregation [21, 32] and preference learning [8, 10, 14], both for general entities and in the Multi-Agent context. Allen [1] has extended the framework to finite attribute domains and indifference, while others [2, 22] have tackled the problem of deriving total orders from a given CP-net.

Yet, those preference-based approaches represent orders over environmental states, while I need to describe orders over action spaces, depending on the value of the environmental attributes. Although these approaches cannot (as illustrated in Section 4.1.1) lead to accurate results in case of a discrepancy between observed and intended behaviour, they still have some interesting implications for the present scenario: First, they show how dependencies between attributes can be used to achieve a more compact and exploitable data structure. Second, the process of deriving knowledge about agent behaviour from observing them is similar (when preferences are not already assumed to be known, as in [11]), such that the use of an analogous structure seems a reasonable next step for my Governance approach.

The self-adaptivity and self-organisation properties of Multi-Agent Systems have been seen as related to Organic Computing Systems by several researchers [20]. The GMAS approach targets the conflicts stemming from differing agent goals and from lack of cooperation by introducing a mediating Governance instance. A similar line of thought was established in [41] in the context of self-organisation and the emergence of cooperation.

4.5 Application for Evaluation

The domain chosen for Example 2 lends itself on several levels to examination as an MAS with system objectives and subsequent need for governance: The agents can have conflicting goals and only express them by acting within the system, there are dependencies between agent actions, and there are undoubtedly undesirable states which should be avoided even if this requires restricting the agents. However, it lacks two more criteria which make an interesting case for an online self-learning Governance, especially as a proof-of-concept for the contribution of the PhD project—Safety-criticality and real-time requirements. Those criteria are satisfied by another application domain: Autonomous vehicles.

The current baseline for designing autonomous cars is that they have to obey the (static) local traffic rules, which includes the ability to detect anomalies and dangers and react accordingly. These regulations are identical for all road users and do not, in general, take into account any specific agent goals. As a consequence, avoiding traffic jams or shortages of parking space can only be addressed globally or via explicit human intervention.

I claim that a self-learning Governance which is given a set of objectives for an autonomous traffic scenario can achieve this to a high extent in an ad-hoc fashion while ensuring compliance with basic safety rules.

Since similar scenarios have been examined in related work, it should be possible to establish a well-defined baseline against which the performance of the GMAS approach can be measured.

4.6 Conclusion

In Multi-Agent research, there is a large gap between agent-centric and system-focused (or governance-focused) learning methods. While individual agents experience a lot of attention from the Game Theory, Logic and Machine Learning communities, governance (both centralised and distributed) leads more of a niche existence, and oftentimes the prerequisites regarding agent behaviour are very specific.

I am aiming towards closing this gap and advancing the area of Governed Multi-Agent Systems such that both effective and minimally restrictive governance becomes available for large and currently uncontrollable systems. To achieve this, formal models and efficient data structures are just as important as governance algorithms which can deal autonomously with incomplete information and unknown, ever-changing agent strategies.

References

1. Allen, T.E.: CP-nets with indifference. In: 2013 51st annual allerton conference on communication, control, and computing (allerton). pp. 1488–1495 (2013)
2. Aydogan, R., Baarslag, T., Hindriks, K., Jonker, C., Yolum, P.: Heuristic-Based Approaches for CP-Nets in Negotiation. In: Studies in Computational Intelligence, vol. 435, pp. 113–123 (Jan 2013), journal Abbreviation: Studies in Computational Intelligence
3. Bade, S.: Nash Equilibrium in Games with Incomplete Preferences. Economic Theory 26(2), 309–332 (2005), www.jstor.org/stable/25055952, publisher: Springer
4. Balke, T., da Costa Pereira, C., Dignum, F., Lorini, E., Rotolo, A., Vasconcelos, W., Villata, S.: Norms in MAS: Definitions and Related Concepts (Jan 2013), pages: 31
5. Boella, G., van der Torre, L., Verhagen, H.: Introduction to normative multiagent systems. Computational & Mathematical Organization Theory 12(2), 71–79 (Oct 2006), https://doi.org/10.1007/s10588-006-9537-7
6. Boutilier, C., Brafman, R.I., Domshlak, C., Hoos, H.H., Poole, D.: CP-Nets: A tool for representing and reasoning with conditional ceteris paribus preference statements. J. Artif. Int. Res. 21(1), 135–191 (Feb 2004)
7. Bulling, N., Dastani, M.: Norm-based Mechanism Design. Artif. Intell. 239(C), 97–142 (Oct 2016), https://doi.org/10.1016/j.artint.2016.07.001
8. Chevaleyre, Y., Koriche, F., Mengin, J., Zanuttini, B.: Learning Ordinal Preferences on Multiattribute Domains: the Case of CP-Nets. Preference Learning (Jan 2011)
9. Claus, C., Boutilier, C.: The Dynamics of Reinforcement Learning in Cooperative Multiagent Systems. In: Proceedings of the Fifteenth National/Tenth Conference on Artificial Intelligence/Innovative Applications of Artificial Intelligence. pp. 746–752. AAAI '98/IAAI '98, American Association for Artificial Intelligence, Menlo Park, CA, USA (1998), http://dl.acm.org/citation.cfm?id=295240.295800
10. Cornelio, C., Goldsmith, J., Mattei, N., Rossi, F., Venable, K.B.: Updates and Uncertainty in CP-Nets. In: Cranefield, S., Nayak, A. (eds.) AI 2013: Advances in Artificial Intelligence. pp. 301–312. Springer International Publishing, Cham (2013)
11. Dell'Anna, D., Dastani, M., Dalpiaz, F.: Runtime Revision of Norms and Sanctions Based on Agent Preferences. In: Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems. pp. 1609–1617. AAMAS '19, International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC (2019), event-place: Montreal QC, Canada

12. García-Camino, A., Rodríguez-Aguilar, J., Sierra, C., Vasconcelos, W.: A rule-based approach to norm-oriented programming of electronic institutions. SIGecom Exchanges 5 (Jan 2006)
13. Gärdenfors, P.: Belief Revision: an introduction. In: Belief Revision (May 1992), journal Abbreviation: Belief Revision
14. Guerin, J.T., Allen, T.E., Goldsmith, J.: Learning CP-net Preferences Online from User Queries. In: Perny, P., Pirlot, M., Tsoukiàs, A. (eds.) Algorithmic Decision Theory. pp. 208–220. Springer Berlin Heidelberg, Berlin, Heidelberg (2013)
15. Gutierrez, J., Perelli, G., Wooldridge, M.: Imperfect information in reactive modules games. Information and Computation 261, 650 – 675 (2018)
16. Hernandez-Leal, P., Kartal, B., Taylor, M.: A survey and critique of multiagent deep reinforcement learning. Autonomous Agents and Multi-Agent Systems (Oct 2019)
17. Hoen, P.J.t., Tuyls, K., Panait, L., Luke, S., La Poutré, J.A.: An Overview of Cooperative and Competitive Multiagent Learning. In: Tuyls, K., Hoen, P.J., Verbeeck, K., Sen, S. (eds.) Learning and Adaption in Multi-Agent Systems. pp. 1–46. Lecture Notes in Computer Science, Springer, Berlin, Heidelberg (2006)
18. Jennings, N.R., Wooldridge, M.J.: Agent technology: foundations, applications, and markets. Springer Science & Business Media (2012)
19. Kantert, J., Edenhofer, S., Tomforde, S., Hähner, J., Müller-Schloer, C.: Normative control: Controlling open distributed systems with autonomous entities. In: Trustworthy Open Self-Organising Systems, pp. 89–126 (2016)
20. Krupitzer, C., Breitbach, M., Roth, F.M., VanSyckel, S., Schiele, G., Becker, C.: A survey on engineering approaches for self-adaptive systems (extended version) (2018), https://madoc.bib.uni-mannheim.de/44034/
21. Kyaw, H., Ghosh, S., Verbrugge, R.: Multi-player multi-issue negotiation with mediator using CP-nets. ICAART 2013 - Proceedings of the 5th International Conference on Agents and Artificial Intelligence 1, 99–108 (Jan 2013)
22. Lang, J., Mengin, J.: The complexity of learning separable ceteris paribus preferences. In: Proceedings of the 21st international jont conference on artifical intelligence. pp. 848–853. IJCAI'09, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (2009)
23. List, C.: Social choice theory. In: Zalta, E.N. (ed.) The Stanford encyclopedia of philosophy. Metaphysics Research Lab, Stanford University, winter 2013 edn. (2013), https://plato.stanford.edu/archives/win2013/entries/social-choice/
24. Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A., Veness, J., Bellemare, M., Graves, A., Riedmiller, M., Fidjeland, A., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., Hassabis, D.: Human-level control through deep reinforcement learning. Nature 518, 529–33 (Feb 2015)
25. Morales, J.: On-line norm synthesis for open Multi-Agent systems. Ph.D. thesis, Universitat de Barcelona (2016)
26. Morris-Martin, A., De Vos, M., Padget, J.: Norm emergence in multiagent systems: a viewpoint paper. Autonomous Agents and Multi-Agent Systems 33(6), 706–749 (Nov 2019), https://doi.org/10.1007/s10458-019-09422-0
27. Müller-Schloer, C., Tomforde, S.: Organic Computing - Technical Systems for Survival in the Real World. Birkhäuser (2017)
28. Nowé, A., Vrancx, P., De Hauwere, Y.M.: Game Theory and Multi-agent Reinforcement Learning. In: Wiering, M., van Otterlo, M. (eds.) Reinforcement Learning: State-of-the-Art, pp. 441–470. Springer Berlin Heidelberg, Berlin, Heidelberg (2012), https://doi.org/10.1007/978-3-642-27645-3_14

29. Perelli, G.: Enforcing Equilibria in Multi-Agent Systems. In: Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems. pp. 188–196. AAMAS '19, International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC (2019), event-place: Montreal QC, Canada
30. Pernpeintner, M.: Toward a self-learning governance loop for competitive multi-attribute mas (submitted)
31. Rizk, Y., Awad, M., Tunstel, E.: Decision Making in Multi-Agent Systems: A Survey. IEEE Transactions on Cognitive and Developmental Systems PP, 1–1 (May 2018)
32. Rossi, F., Venable, K., Walsh, T.: mCP Nets: Representing and Reasoning with Preferences of Multiple Agents. (Jan 2004), journal Abbreviation: Proceedings of the National Conference on Artificial Intelligence Pages: 734 Publication Title: Proceedings of the National Conference on Artificial Intelligence
33. Rotolo, A.: Norm compliance of rule-based cognitive agents. pp. 2716–2721. IJCAI International Joint Conference on Artificial Intelligence (Jan 2011)
34. Rotolo, A., van der Torre, L.: Rules, Agents and Norms: Guidelines for Rule-Based Normative Multi-Agent Systems. In: Bassiliades, N., Governatori, G., Paschke, A. (eds.) Rule-Based Reasoning, Programming, and Applications. pp. 52–66. Springer Berlin Heidelberg, Berlin, Heidelberg (2011)
35. Russell, S., Norvig, P.: Artificial intelligence: A modern approach. Prentice Hall Press, USA, 3rd edn. (2009)
36. Shoham, Y., Powers, R., Grenager, T.: Multi-Agent Reinforcement Learning: a critical survey (Jun 2003)
37. Shoham, Y., Tennenholtz, M.: On social laws for artificial agent societies: off-line design. Artificial Intelligence 73(1), 231 – 252 (1995), http://www.sciencedirect.com/science/article/pii/000437029400007N
38. Stirling, W.C., Felin, T.: Game theory, conditional preferences, and social influence. PLOS ONE 8(2), 1–11 (Feb 2013), https://doi.org/10.1371/journal.pone.0056751
39. Watkins, C.: Learning From Delayed Rewards (Jan 1989)
40. Weyns, D., Michel, F.: Agent environments for multi-agent systems – a research roadmap. In: Weyns, D., Michel, F. (eds.) Agent environments for multi-agent systems IV. pp. 3–21. Springer International Publishing, Cham (2015)
41. Wolf, T.D., Holvoet, T.: Emergence and self-organisation: a statement of similarities and differences (2004)