Supporting the Reasoning about Environmental Consequences of Institutional Actions

Rafhael R. Cunha[1,2] [0000−0003−3233−5158], Jomi F. Hübner[2] [0000−0001−9355−822X], and Maiquel de Brito[3] [0000−0003−4650−7416]

[1] Federal Institute of Education, Science and Technology of Rio Grande do Sul (IFRS), Campus Vacaria

[2] Automation and Systems Department, Federal University of Santa Catarina, Florianópolis, Brazil

rafhael.cunha@posgrad.ufsc.br

jomi.hubner@ufsc.br

[3] Control, Automation, and Computation Engineering Department, Federal University of Santa Catarina, Blumenau, Brazil

maiquel.b@ufsc.br

Abstract. In this paper we are considering multi-agent systems (MAS) with agents that have both goals and anti-goals. Goals represent environment states that agents want to achieve and anti-goals represent environment states they want to avoid. To achieve their goals, agents perform some actions that may have institutional consequences. Which could potentially change the environment towards as a counter effect. Since these consequences are institutional, they should be explicitly specified so that agents are able take them into consideration in their decision process. However, existing models of artificial institutions do not consider such consequences. Considering this problem, this paper proposes to extend the institutional specification making explicit the implications of the institutional actions in the environment. The proposal is presented, discussed and implemented using the JaCaMo framework, highlighting its advantages for agents while reasoning about the consequences of their action both in the institution and the environment.

Keywords: Purposes · Status-functions · Artificial institutions.

1 Introduction

The achievement of the goals of an agent may depend on some status assigned to the actions that it performs instead of depending on the actions themselves. Consider a scenario where an agent called *sBob* has the goal of *conquering a new territory*. The agent knows from some available guidelines that the goal is achieved by performing a digital action (e.g. sending a message, posting on a webservice) that has the status (or *counts as*) *commanding an attack*. This action is supposed to produce in the environment the effects corresponding to such status (e.g. destroying buildings, killing opponents, etc.). However, these

effects may not be explicit to the agent. It can choose the action to perform based solely on its status.

Inspired by human societies, some works propose models and tools to manage the assignment of *statuses* to the elements involved in the Multi-Agent System (MAS) [15]. In this paper, the element of the system in charge of managing the assignment of status is called *institution*. Through the institution, agents may have the status of *soldiers*, while some of their digital actions may have the status of *commanding an attack*. These works focus on assigning status to the elements that compose the MAS in a process called constitution. However, they do not address the effects in the environment[4] of such statuses. For example, a digital action with the assigned status of *commanding an attack* can trigger a series of consequences in the environment such as *killing a soldier from allied base*, *killing innocent people*, etc. that may be unknown/unwanted to agents and that would not happen if the action did not have the status.

There are some drawbacks of not specifying the consequences in the environment of actions that have a status (see more in [9]). This work focuses on actions whose status leads to a goal achievement but whose effects in the environment are undesirable for the agent. For example, consider an institutional specification stating that *sending a broadcast message has the status of commanding an attack*. In this case, *sBob* can use this specification to discover how to achieve its goal, i.e., by broadcasting a message. However, if the institutional specification does not express the effects in the environment of commanding an attack, *sBob* can not rely on this specification to discover the consequences of broadcasting the message. If *sBob* has the principle of not killing a soldier from the allied base but commanding an attack can make this consequence possible, *sBob* may violate its principle if not aware of these consequences.

Regarding these issues, the main contribution of this paper is a proposal of a mechanism that allows agents to discover what are the consequences in the environment of performing an action that has a status. This proposal is inspired by the Construction of the social reality by John Searle [23, 22] theory that seems to be fundamental for comprehending the social reality.

This paper is organized as follows: Sect. 2 introduces the main background concepts necessary to understanding our proposal and its position in the literature. It includes philosophical theory and related works. Sect. 3 presents the proposed model, its definitions and functions and algorithms to use of the model. Sect. 4 illustrates how the use of artificial institutions and purposes facilitates the development of agents capable of reasoning about the implications of status actions in the environment. Finally, Sect. 5 presents some conclusions about this work and suggests future works.

[4] In this paper, *environment* refers to the set of physical and digital resources which the agents perceive and act upon [26].

2 Artificial institutions

The problem described in the introduction is rooted in the fact that concrete elements of MAS may have statuses that are not necessarily related to their design features. In MAS, these statuses are managed by Artificial Institutions. Artificial Institutions are inspired by John Searle's theory [23, 22], which claims that the social reality where human beings are immersed arises from the concrete world (i.e., the environment) based on some elements, including status-functions and constitutive rules. *Status-functions* are *status* that assign *functions* to the concrete elements [23, 22]. These functions cannot be explained through their physical virtues. For example, the status *buyer* assigns to an agent some functions such as perform payments, take loans, etc. *Constitutive rules* specify the assignment of status-functions to concrete elements with the following formula: X count-as Y in C. For example, a piece of paper count-as money in a bank, where X represents the concrete element, Y the status-function, and C the context where that attribute is valid. The attribution of status-functions through constitutive rules to environment elements is called constitution and creates institutional facts. The set of institutional facts gives rise to institutions [22]. Artificial Institutions (or simply *institutions*) are the component of the MAS that is responsible for defining the conditions for an agent to become a *buyer*, or an action to become a *payment* [23, 22].

Works on Artificial Institutions are usually inspired by the theory of John Searle [23, 22]. Some works present functional approaches, relating brute facts to normative states (e.g., a given action counts as a violation of a norm). These works do not address ontological issues, and, therefore, it becomes even more difficult to support the meaning of abstract concepts present in the institutional reality. Other works have ontological approaches, where brute facts are related to concepts used in the specification of norms (e.g., sending a message counts as a bid in an auction). However, these works have some limitations.

Some approaches allow the agents to reason about the constitutive rules [8, 11, 10, 6, 25, 1]. However, generally the *status-function* (Y) is a label assigned to the concrete element (X) that is used in the specification of the regulative norms. Therefore, Y does not seem to have any other purpose than to serve as a basis for the specification of stable regulative norms [24, 1]. Some exceptions are (i) in [12, 13, 11, 14] where Y represents a class formed with some properties as roles responsible for executing actions, time to execute them, condition for execution, etc.; (ii) in [24] where Y is a general concept, and X is a sub-concept that can be used to explain Y. Although the exceptions contain more information than just a label in the Y element, these data are somehow associated with regulative norms.

In short, existing works in artificial institutions are mainly concerned with specifying and managing the constitution. However, the constitution is based on facts occurring in the environment that may even produce further environmental consequences. While the constitution is explicit, *it is implicit in these works the environmental states that can be reached because an action constitutes a status-function.* In the previous example, while the constitutive rule specifies how to

constitute *commanding an attack*, the effects in the environment of commanding an attack are not explicit. Some agent cannot rely on the institutional specification to evaluate the effects in the environment of achieving a goal that depends on the constitution of a status-function. Designers make this association between the constitution of a status-function and its environmental consequences in an ad-hoc manner. The main disadvantage of an ad-hoc association is that the agent works only in scenarios foreseen by the developer.

The limitation discussed indicates the need to develop a model that explains the purposes of status-functions belonging to institutional reality. Aguilar et al. [21] corroborate this conclusion by stating that institutions have not yet considered how to help agents in decision-making, helping them to achieve their own goals. The modeling of purposes of status-functions, described in the next section, is a step to fill this open gap.

3 The purposes of Status-Functions

The mentioned issues are associated with the relationship between constitutions of status-functions and their consequences in the environment. While works on MAS ignore these relations, Searle addresses them under the notion of *Purpose* [23, 22]. Functions related to statuses are called agentive functions because they are assigned from the practical interests of agents [23, p.20]. These practical interests of agents are called *purposes* [22, p.58]. Thus, the purposes point to the consequences in the environment of the constitution of status-functions that are aligned with the agents' interests. For example, someone has a goal of *inhabiting a piece of land* when he broadcasts a message that institutionally is considered as *commanding an attack*. In this example, *inhabiting a piece of land* represents a state of the world that is pointed by a purpose. This state is enabled (and will probably happen) when the status-function *commanding an attack* is constituted. The states must reflect the interest of the agents involved in that context. Moreover, the agents involved in the interaction should have a common understanding of these facts and purposes and consider them in their deliberation. Otherwise, none of them achieve their social goal[5].

The essential elements of the proposed model are *agents*, *states*, *institutions*, and *purposes*, depicted in the Figure 1. *Agents* are autonomous entities that pursue their goals in the MAS [28]. The literature presents several definitions of *goal* that are different but complementary to each other (see more in [3, 27, 16, 20, 17, 18]). In this work, *goals* are something that agents aim to achieve (e.g. a certain state, the performance of an action. According to Aydemir, et.al [2], *anti-goal* is an undesired circumstance of the system. In this work, anti-goal represents states that the agent does not wish to reach for ethical reasons, particular values, prohibition by some regulative norm, etc. Moreover, agents can perform actions that trigger events in the MAS. If this action produces events that may constitute some status-function, this action is an *institutional action*. *States* are formed by

[5] In this paper, a social goal is an goal that depends on other agents acting on the system.

one or more properties that describe the characteristics of the system at some point of its execution [7].

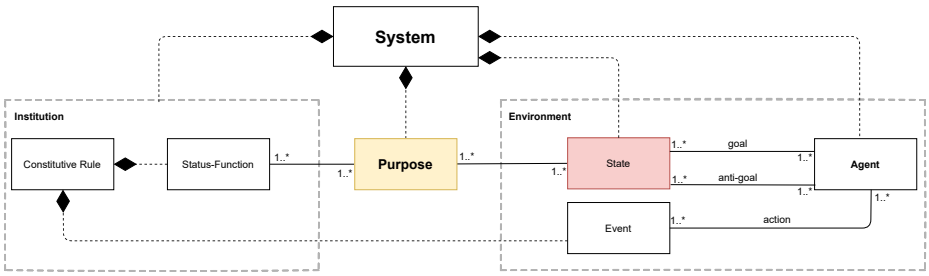

Fig. 1. Overview of the model.

Institutions provide the social interpretation of the environmental elements of the MAS as usually proposed in the literature. This social interpretation occurs through the interpretation of constitutive rules that assign status to environmental elements, as described in Section 2. It is beyond the scope of this paper to propose a model of artificial institution. Rather, it considers this general notion of the institution as the entity that constitutes status-functions, that is adopted by several models in the field of MAS.

While *agents*, *states* and *institutions* are known concepts, *purposes* are introduced in this model. The functions associated with status-functions can satisfy the practical interests of agents. From the institution's perspective, these interests are called *Purposes*. From the agents' perspective, these interests are their goals or anti-goals. Then, we claim that (i) *the goals or anti-goals of the agents match with the purposes of the status-functions* and (ii) *goals, anti-goals and purposes point to environmental states related to the status-functions*. For example, in the war scenario, an agent that performs an action that counts as *commanding an attack* triggers intermediate events that bring the system to states such as *conquer a new territory* (i.e., the agent goal) or *killing a soldier from the allied base* (i.e., the agent anti-goal). The intermediate events (e.g. shoot someone) between the constitution of the status-functions and the environmental states reached are ignored in our proposal, since we consider that the agent is only interested in the states that can be reached after the status-functions is constituted.

Shortly, this model provides two relationships: (i) between purposes and status-functions and (ii) between purposes and agent goals and anti-goals. Thus, if (i) there is a constitutive rule specifying how a status-function is constituted, (ii) a purpose associated with that status-functions, and (iii) an agent that has a goal or anti-goal that matches with the states pointed to by the purpose, then it is explicit how the agent should act to achieve its goal or avoid an anti-goal. In the previous example, *sBob* can know that if it constitutes the status-function *commanding an attack* to satisfy its goal of *conquering a new territory*, some

other states will be reached such as *killing a soldier from the allied base*, *killing innocent people*, which may be undesirable to the agent.

3.1 Definitions

This section formally[6] describes the model by specifying (i) the purposes associated with the status-functions and (ii) the purposes associated with the consequences in the environment of constituting status-functions. These consequences are states of the world that agents want to reach or prefer to avoid. Although the concept of purpose is independent, it is used in conjunction with the states, agents and institutions that make up the MAS.

Definitions 1 to 5 represent the MAS states, events, agents, agents goals and anti-goals, and the relationships that exist between these concepts. These definitions express the environmental elements that belong to the MAS (expressed in the Environment rectangle in the Figure 1). Definitions 6 and 7 are imported from the Situated Artificial Institution (SAI) model [5, 10] and represent the elements that make up the institution and its connection with the environmental elements (expressed in the Institution rectangle in the Figure 1). The definitions 8 to 11 represent the purposes and the relationships that exist between them and the institution and between purpose and the states of the world that agents wish to achieve or avoid (expressed in the Purpose rectangle and its relations in the Figure 1).

Definition 1 (States). *Properties are characteristics of the system at some point of its execution. The set of all properties that the system can present is represented by* $\mathcal{T}$*. The state of the system at some point of its execution is the set of all the standing properties.* $\mathcal{S} = 2^{\mathcal{T}}$ *is the set of all the possible states of the MAS. For example, the sets* $\mathsf{s1} = \{territory\ conquered\}$ *and* $\mathsf{s2} = \{killed\ from\ allied\ base\}$ *define states that exist in the MAS, where* $\mathsf{s1} \in \mathcal{S}$ *and* $\mathsf{s2} \in \mathcal{S}$*.*

Definition 2 (Events). *Event is an instantaneous occurrence within the system [7]. Events may be both triggered by actions of the agents (e.g. sending of a message) and spontaneously produced by some non autonomous element (e.g. a clock tick). The set of all events that may happen in the system is represented by* $\mathcal{E}$*. Each event is represented by an identifier. For example, the set* $\mathcal{E} = \{broadcast\ a\ message\}$ *defines the event that can happen in the MAS.*

Definition 3 (Agents). *The set of all agents that can act in the MAS is represented by* $\mathcal{A}$*. Each agent is represented by an identifier. For example, the set* $\mathcal{A} = \{sBob\}$ *defines the agent that exists in the MAS.*

Definition 4 (Relationship between Agents and their goals). *In this work, agents goals are states of the world that agents desire to reach*[7]*. The set of*

[6] We formalize the model to make it more accurate and facilitate the development of algorithms that can be used to improve the agents' decision process.

[7] We focus on declarative goals (i.e., goals that describe desirable situations) because we are interested in the effects of the constitution of status-functions that may even

the goals of the agents acting in the system is given by $\mathcal{G} \subseteq \mathcal{A} \times \mathcal{S}$*. For example, the pair* $\langle sBob, territory\ conquered \rangle \in \mathcal{G}$ *means that the agent* sBob *has the goal territory conquered.*

Definition 5 (Relationship between Agents and their anti-goals). *Anti-goals are states in the MAS that agents desire to avoid. The set of the anti-goals of the agents is given by* $\overline{G} \subseteq \mathcal{A} \times \mathcal{S}$*. For example, the pair* $\langle sBob, soldier\ killed\ from\ allied\ base \rangle \in \overline{G}$ *means that the agent* sBob *has the anti-goal "soldier killed from allied base". From a general point of view, there is no difference between an anti-goal and the denial of a goal (the negation of a goal). However, to avoid the addition of negated goals in the model, we opted to have explicit anti-goals. The intersection between agent goal and anti-goal should be empty* $(\mathcal{G} \cap \overline{G} = \emptyset)$*.*

Definition 6 (Status-Functions). *A* status *is an identifier that assigns to the environmental elements an accepted position, especially in a social group. It allows the environmental elements to perform* functions *(associated with the status) that cannot be explained through its physical structure [22, p.07]. For simplicity, in this formalization we only consider statuses assigned to events. The set of all the event-status-functions of an institution is represented by* $\mathcal{F}$*. For example, the set* $\mathsf{f} = \{command\ an\ attack\}$ *defines a status that exists in the MAS, where* $\mathsf{f} \subseteq \mathcal{F}$*.*

Definition 7 (Constitutive rules). *Constitutive rules specify the constitution of status-functions from environmental elements. Searle proposes to express these rules as* X count-as Y in C*, explained in Section 2. Since the process of constitution is beyond the scope of this paper, the element C can be ignored. For simplicity, a constitutive rule is hereinafter expressed as* X count-as Y*. The set of all constitutive rules of an institution is represented by* $\mathcal{C}$*. A constitutive rule* $c \subseteq \mathcal{C}$ *is a tuple* $\langle x, y \rangle$*, where* $x \in \mathcal{E}$ *and* $y \in \mathcal{F}$*, meaning that x count-as y. For example, the set* $c = \{\langle broadcast\ a\ message, command\ an\ attack\ \rangle\}$ *defines a constitutive rule related to the scenario.*

Definition 8 (Purposes). *The purposes are related to the agents' practical interests. We assume that the set of all purposes is represented by* $\mathcal{P}$*. Each purpose is represented by an identifier. For example, the set* $\mathcal{P} = \{new\_territory\}$*, define the unique purpose that exists in the MAS.*

Definition 9 (Relationship between status-functions and purposes). *We define that purposes can be satisfied through the constitution of status-functions. Thus, there must be a relationship between these two concepts. This relation is represented by* $\mathcal{F}_P \subseteq \mathcal{F} \times \mathcal{P}$*. For example,* $\{\langle command\ an\ attack,\ new\_territory \rangle\} \in \mathcal{F}_P$ *means that the constitution of the status-function command an attack satisfies the purpose new territory.*

produce further environmental consequences (i.e., new states of the world). There are some other types of goals (e.g. procedural goal) that focus on the execution of the action and therefore are not compatible with the concept of purpose.

Definition 10 (Relationship between purposes and agent's goals and anti-goals). *The relationship between purpose and agent goal and anti-goal considers that a purpose point to one or more states in the MAS that matches the agents goals and anti-goals. The relationship* $\mathcal{G}_P$ *is a tuple* $\langle \mathsf{p}, \mathsf{a}_{gag} \rangle$ *where* $\mathsf{p} \in \mathcal{P}$ *and* $\mathsf{a}_{gag} \in 2^{\mathcal{G} \cup \overline{G}}$*. For example, the set* $\mathcal{G}_P = \{\langle new\_territory, \{territory\ conquered\}, \{soldier\ killed\ from\ allied\ base\}\rangle\}$ *defines the relation that exists between the purpose and the states of the world that it points that match with agents' goal or anti-goal.*

Definition 11 (Model). *The model is a tuple* $\langle \mathcal{S}, \mathcal{E}, \mathcal{A}, \mathcal{A}_{GA}, \mathcal{F}, \mathcal{C}, \mathcal{P}, \mathcal{F}_P, \mathcal{G}_P \rangle$*, where* $\mathcal{S}$ *is the set of states that may be maintained in the MAS,* $\mathcal{E}$ *is the set of events happen that may happen in the MAS,* $\mathcal{A}$ *is the set of agents that can act in the MAS,* $\mathcal{A}_{GA}$ *is the set of goals and anti-goals of agents (i.e.,* $\mathcal{A}_{GA} = \mathcal{G} \cup \overline{G}$*),* $\mathcal{F}$ *is the set of status-functions,* $\mathcal{C}$ *is the set of constitutive-rules that may exists in the MAS,* $\mathcal{P}$ *is the set of purposes,* $\mathcal{F}_P$ *is set that expresses the relationship between the* $\mathcal{F}$ *and* $\mathcal{P}$ *sets and* $\mathcal{G}_P$ *is the set that represents the relationship between* $\mathcal{P}$ *and* $\mathcal{A}_{GA}$*.*

3.2 Functions and algorithms

In this section we formalize some functions that can be used by an agent to discover the environmental effects of performing an institutional action. For that, we need the status-functions related to the events produced by an action (Definition 14), the purposes of these status-functions (Definition 12), and the states of these purposes (Definition 13). In the example of this paper, *sBob* knows by doing *broadcast a message* that it satisfies its goal. With the proposed functions, it can discover that this action has other consequences (e.g., someone being killed) which are among its anti-goals. It may thus avoid that action to achieve its goal.

Definition 12 (Mapping status-functions to purposes). *Given a set* $\mathcal{F}$ *of status-functions and a set* $\mathcal{P}$ *of purposes, the set of purposes that are enabled when a status-function is constituted is given by the function* $fp : \mathcal{F} \rightarrow 2^{\mathcal{P}}$ *s.t.* $fp(\mathsf{f}) = \{\mathsf{p} \mid \langle \mathsf{f}, \mathsf{p} \rangle \in \mathcal{F}_P\}$*.*

For example, if $\mathcal{F}_P = \{\langle command\ an\ attack,\ new\_territory \rangle\}$, then $fp(command\ an\ attack) = \{new\_territory\}$.

Definition 13 (Mapping purposes to states). *Given a set* $\mathcal{P}$ *of purposes and a set* $\mathcal{A}_{GA}$ *(*$\mathcal{A}_{GA} = \mathcal{G} \cup \overline{G}$*) of agents goals and anti-goals, the set of agents goals and anti-goals that are pointed by a purpose is given by the function* fsw *:* $\mathcal{P} \rightarrow 2^{\mathcal{A}_{GA}}$ *s.t.* $fsw(p) = \{\mathsf{a}_{ga} \mid \langle \mathsf{p}, \mathsf{a}_{ga} \rangle \in \mathcal{G}_P\}$*.*

For example, if $\mathcal{G}_P = \{\langle new\_territory, \{territory\ conquered\}, \{soldier\ killed\ from\ allied\ base\}\rangle\}$, then $fsw(new\_territory) = \{\{territory\ conquered\}, \{soldier\ killed\ from\ allied\ base\}\}$.

Definition 14 (Mapping events to status-functions). *Given a set $\mathcal{F}$ of status-functions and a set of events $\mathcal{E}$, the status-functions that are constituted by an event are given by the function $fc : \mathcal{E} \rightarrow 2^{\mathcal{F}}$ s.t. $fc(\mathsf{e}) = \{\mathsf{f} \mid \langle \mathsf{e}, \mathsf{f} \rangle \in \mathcal{C}\}$.*

For example, if $\mathcal{C} = \{\langle broadcast\ a\ message, command\ an\ attack \rangle\}$, then $fc(broadcast\ a\ message) = \{command\ an\ attack\}$.

Definition 15 (Mapping status-functions to events). *Given a set $\mathcal{F}$ of status-functions and a set of events $\mathcal{E}$, the events that constitute the status-functions are given by the function $fca : \mathcal{F} \rightarrow 2^{\mathcal{E}}$ s.t. $fca(\mathsf{f}) = \{\mathsf{e} \mid \langle \mathsf{e}, \mathsf{f} \rangle \in \mathcal{C}\}$.*

For example, if $\mathcal{C} = \{\langle broadcast\ a\ message, command\ an\ attack \rangle\}$, then $fca(command\ an\ attack) = \{broadcast\ a\ message\}$.

From these functions, the Algorithm 1 can be used by the agent to find out which are the environmental effects if some action is executed in an institutional context. The algorithm can be summarized in some steps: (1) verify whether the action is an institutional action, i.e., it its events constitutes something in the institution (lines 4 and 5), if true, go to the next step, otherwise returns the empty set (line 12); (2) consider all status-functions related to the action (line 6); (3) consider all purposes of such status-functions (line 7); and (4) for each purpose, looks for the states it points to and add them in the answer of the algorithm.

Algorithm 1 Find the effects of an action in the environment

```
1: Input: an action ac
2: Output: the set of possible states after ac
3: s ← {}
4: e ← event produced by action ac
5: if fc(e) ≠ {} then                    ▷ if the event e may constitute a status-functions
6:     for f ∈ fc(e) do                  ▷ f is the set of status-functions that e count-as
7:         for p ∈ fp(f) do              ▷ p is the set of purposes that are associated with f
8:             s ← s ∪ fsw(p)            ▷ add states pointed to by p
9:         end for
10:    end for
11: end if
12: return s
```

To verify if some action can produce some state considered as an anti-goal, we developed algorithm 2. To illustrate it, in the case of *sBob* considering the action *broadcast a message* to achieve some goal, the execution of the algorithm for this action returns *true*, meaning that the action can also produce effects considered as an anti-goal.

Algorithm 2 Verifies whether some action can produce states considered as anti-goals.

1: **Input:** $\overline{G}$, ac
2: **Output:** returns true if ac implies anti-goals and false otherwse
3: se ← algorithm 1(ac) ▷ se is the set of states pointed to by ac
4: **return** $\exists_{\mathsf{ag}\in\overline{G}}$ ag ∈ se ▷ checks whether anti goals are included in se

4 Implementing the purpose model

To illustrate the use of this model, we recall the example introduced at the beginning of this paper: the scenario where *sBob* desires to reach its goal of *territory conquered*. To this end, *sBob* knows that to achieve *territory conquered*, it needs to perform an (institutional) action that count-as *commanding an attack*. From the constitutive rule — *broadcast a message count-as commanding an attack* — it knows that it needs to *broadcast a message* to achieve its goal. The purpose model it is possible to specify that the status-function *commanding an attack* is associated with the purpose *new_territory*, which, on its turn, is associated with a state with the following properties: *territory conquered* and *soldier killed from the allied base*. Thus, *sBob* is now able to reason about the consequences of performing the action *broadcast a message* in the institutional context. Such an institution could include other status-functions but, for simplicity, we focus only on those essential to illustrate the main features of the model proposed in Section 3.

The example is implemented through the components depicted in Figure 2. The agent *sBob* is programmed in Jason [4] and the environment in CArtAgO [19]. To implement the artificial institution, we use an implementation of the Situated Artificial Institution model (SAI) model [10]. It provides means to specify status-functions and constitutive rules and to manage the constitution process. The purpose model is implemented through an ontology encapsulated in a CArtAgO artifact which is accessible to the agents. The query and persistence of data in the ontology are enabled by the MasOntology[8], a set of tools developed in CArtAgO to interact with ontologies[9].

Figure 3 depicts the agent program. Line 1 specifies an anti-goal of *sBob*. *sBob* goal can be achieved by the plan illustrated in lines 3–11. This plan creates subgoals `alg1` and `alg2` that can be achieved by plans in lines 13–22, which are the Jason implementation of Algorithms 1 and 2. Regarding the plan for `alg1`, if the `Action` does not constitute a status-function, the `States` are empty (line 17). Otherwise, some operations are used to retrieve the list of `States` related to the action in lines 14 and 15. Regarding the plan for `alg2`, it simply gets the list of states from `alg1` and tests if some anti-goal is member of this list. The result is unified with variable `R`. The value of `R` is then used to decide whether to

[8] https://github.com/smart-pucrs/MasOntology
[9] An initial implementation of this platform can be found in https://github.com/rafhaelrc/psf_model.

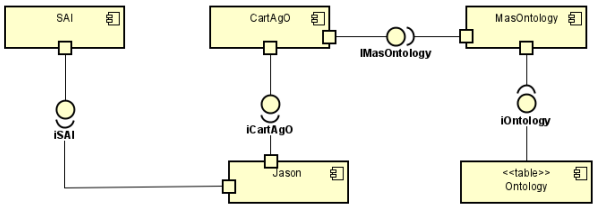

Fig. 2. Component diagram with the systems used to compose the example.

```
1  anti_goal(kill_soldier_from_allied_base).
2
3  +!territory_conquered
4      <- // broadcasting_a_message is the action that achieves the goal territory_conquered
5         !alg2(broadcasting_a_message, R);
6         if(R){
7          .fail;
8         }
9         else{
10         broadcasting_a_message;
11        }.
12
13 +!alg1(Action,States) : constitutive_rule(Action,Status_Functions,_,_)
14     <- getPurposeOfStatusFunctions(Status_Functions, NamePurposes);
15        getStatesOfPurposes(NamePurposes, States).
16
17 +!alg1(Action,[]). // if there is no constitutive rule
18
19 +! alg2(Action, R)
20     <-
21        !alg1(Action, States);
22        R = (anti_goal(AG) & .member(Ag,States)).
```

Fig. 3. Plan of the agent *sBob*.

execute `broadcasting_a_message`, if `R` is `false` it means that the action does not promote some of the agent anti-goals.

The code snippet depicted in Figure 3 illustrates how the algorithms and the model proposed in this work can be used by the agent to check if the action to be performed can constitute a status-functions and enable new states in the system and verify if these new states are unwanted by the agent. We can notice that the code from lines 6 to 11 are just an example of how the proposed model and algorithms can be used. Of course, more complex solutions could be developed for other applications.

5 Conclusions and future work

The problem motivating this paper is some difficulty for agents to reason about the consequences in the environment when performing an action that has an institutional interpretation (i.e., it has a status-function). To help agents with this issue, we introduce the notion of *purpose* in artificial institutions. Purposes connect two concepts: *status-functions* in the institutional side and *goals and anti-goals* in the agent side. While status-functions represent how the environment changes the institution, purposes represent how the institution can potentially

change the environment. From an agent perspective, their goals and anti-goals are also considered in the proposal: purposes point to states of the world that are of interest to the agents. Thus, the model connects institutional facts with the interests of the agents.

The main advantage of purposes in MAS regards the agents. We have an improvement in agent decision-making, since it has more information available to help it to decide whether to achieve its goals or avoid its anti-goals. With the proposed model, agents can access and reason about the consequences of institutional actions and adapt themselves to different scenarios. They can notice that (a) some purposes point to states that are similar to their interests and therefore useful to reach their goals or (b) avoid these purposes because they point to states that are similar to their anti-goals. In both cases, the agent has more information while deciding whether a particular action will help it or not. This kind of reasoning is important for advances in agents autonomy [21].

As future work, we plan to explore additional theoretical aspects related to the proposal, such as (i) investigations about how other proposed institutional abstractions (e.g. social functions) fit on the model, and (ii) check if the purposes related to status must be further detailed. We plan to also address more practical points such as (i) the modeling of a status-functions purposes based on a real scenario, (ii) the implementation of the proposal in a computer system (iii) its integration in an computational model that implements the constitution of status-functions in an MAS platform and (iv) evaluate the application of the model in scenarios that involve ethical reasoning of agents.

Acknowledgments

This study was supported by the Federal Institute of Education, Science and Technology of Rio Grande do Sul (IFRS). We thank the reviewers for the valuable contributions that allowed this work to evolve.

References

1. Aldewereld, H., Álvarez-Napagao, S., Dignum, F., Vázquez-Salceda, J.: Making norms concrete. In: Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: volume 1-Volume 1. pp. 807–814. International Foundation for Autonomous Agents and Multiagent Systems (2010)
2. Aydemir, F.B., Giorgini, P., Mylopoulos, J.: Multi-objective risk analysis with goal models. In: 2016 IEEE Tenth International Conference on Research Challenges in Information Science (RCIS). pp. 1–10. IEEE (2016)
3. Boissier, O., Bordini, R.H., Hubner, J., Ricci, A.: Multi-agent oriented programming: programming multi-agent systems using JaCaMo. MIT Press (2020)
4. Bordini, R.H., Hübner, J.F., Wooldridge, M.: Programming multi-agent systems in AgentSpeak using Jason, vol. 8. John Wiley & Sons (2007)
5. Brito, M.d., et al.: A model of institucional reality supporting the regulation in artificial institutions. Ph.D. thesis, Universidade Federal de Santa Catarina (2016)

6. Cardoso, H.L., Oliveira, E.: Institutional Reality and Norms: Specifying and Monitoring Agent Organizations. International Journal of Cooperative Information Systems **16**(01), 67–95 (2007). https://doi.org/10.1142/s0218843007001573
7. Cassandras, C.G., Lafortune, S.: Introduction to discrete event systems. Springer (2008)
8. Cliffe, O., De Vos, M., Padget, J.: Specifying and reasoning about multiple institutions. In: International Workshop on Coordination, Organizations, Institutions, and Norms in Agent Systems. pp. 67–85. Springer (2006)
9. Cunha, R.R., Hübner, J.F., de Brito, M.: Coupling purposes with status-functions in artificial institutions. arXiv preprint arXiv:2105.00090 (2021)
10. De Brito, M., Hübner, J.F., Boissier, O.: Situated artificial institutions: stability, consistency, and flexibility in the regulation of agent societies. Autonomous Agents and Multi-Agent Systems **32**(2), 219–251 (2018)
11. Fornara, N.: Specifying and monitoring obligations in open multiagent systems using semantic web technology. In: Semantic agent systems, pp. 25–45. Springer (2011)
12. Fornara, N., Colombetti, M.: Ontology and time evolution of obligations and prohibitions using semantic web technology. In: International Workshop on Declarative Agent Languages and Technologies. pp. 101–118. Springer (2009)
13. Fornara, N., Colombetti, M.: Representation and monitoring of commitments and norms using owl. AI communications **23**(4), 341–356 (2010)
14. Fornara, N., Tampitsikas, C.: Using owl artificial institutions for dynamically creating open spaces of interaction. In: AT. pp. 281–295 (2012)
15. Fornara, N., Viganò, F., Colombetti, M.: Agent communication and artificial institutions. Autonomous Agents and Multi-Agent Systems **14**(2), 121–142 (2007). https://doi.org/10.1007/s10458-006-0017-8
16. Hindriks, K.V., De Boer, F.S., Van Der Hoek, W., Meyer, J.J.C.: Agent programming with declarative goals. In: International Workshop on Agent Theories, Architectures, and Languages. pp. 228–243. Springer (2000)
17. Hübner, J.F., Bordini, R.H., Wooldridge, M.: Declarative goal patterns for agentspeak. In: Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'06) (2006)
18. Nigam, V., Leite, J.: A dynamic logic programming based system for agents with declarative goals. In: International Workshop on Declarative Agent Languages and Technologies. pp. 174–190. Springer (2006)
19. Ricci, A., Piunti, M., Viroli, M.: Environment programming in multi-agent systems: an artifact-based perspective. Autonomous Agents and Multi-Agent Systems **23**(2), 158–192 (2011)
20. van Riemsdijk, B., van der Hoek, W., Meyer, J.J.C.: Agent programming in dribble: from beliefs to goals using plans. In: Proceedings of the second international joint conference on Autonomous agents and multiagent systems. pp. 393–400 (2003)
21. Rodriguez-Aguilar, J.A., Sierra, C., Arcos, J.L., Lopez-Sanchez, M., Rodriguez, I.: Towards next generation coordination infrastructures. Knowledge Engineering Review **30**(4), 435–453 (2015). https://doi.org/10.1017/S0269888915000090
22. Searle, J.: Making the social world: The structure of human civilization. Oxford University Press (2010)
23. Searle, J.R.: The construction of social reality. Simon and Schuster (1995)
24. Vázquez-Salceda, J., Aldewereld, H., Grossi, D., Dignum, F.: From human regulations to regulated software agents' behavior. Artificial Intelligence and Law **16**(1), 73–87 (2008)

25. Viganò, F., Colombetti, M.: Model Checking Norms and Sanctions in Institutions (ii), 316–329 (2008)
26. Weyns, D., Omicini, A., Odell, J.: Environment as a first class abstraction in multiagent systems. Autonomous agents and multi-agent systems **14**(1), 5–30 (2007)
27. Winikoff, M., Padgham, L., Harland, J., Thangarajah, J.: Declarative and procedural goals in intelligent agent systems. In: International Conference on Principles of Knowledge Representation and Reasoning. Morgan Kaufman (2002)
28. Wooldridge, M.: An introduction to multiagent systems. John Wiley & Sons (2009)