Shared Strategies in Artificial Agent Societies

Amineh Ghorbani[1], Huib Aldewereld[1], Virginia Dignum[1], and Pablo Noriega[2]

[1] Delft University of Technology, Faculty of Technology, Policy and Management, Delft, The Netherlands
{a.ghorbani,H.M.Aldewereld,M.V.Dignum}@tudelft.nl
[2] Artificial Intelligence Research Institute of the Spanish National Scientific Research Council, Barcelona, Spain
pablo@iiia.csic.es

Abstract. A shared strategy is a social concept that refers to a type of behavioural pattern that is followed by a significant number of individuals although it is, prima facie, not associated with an obligation or a prohibition. E. Ostrom has argued in favour of the pertinence of social strategies for institutional design and evolution and proposed a characterization suggestive of formal treatment. However, shared strategies as such have not been explicitly used in the context of regulated MAS in spite of their relevance and their affinity to more standard normative notions, of which a rich tradition exists in MAS research. In this paper, we discuss the notion of shared strategy, characterize its distinguishing features, propose its formalization using a temporal epistemic logic, and explore its potential use in regulated multi-agent systems.

1 Introduction

In the Netherlands, almost all people have dinner around 5:30pm. As a foreigner in that country, it is almost impossible to plan a (working) meeting around this time, which would be a 'normal' time in many other countries. On the other hand, having dinner that early is not an obligation. No one will be offended or would even care if you choose to eat later. In other words, there is no particular goal that everyone must reach following this strategy and therefore, individual disobedience would not have any particular consequence. One other important attribute of such statement is that it is more significant at the collective level rather than individual. In many cases people are not even aware of the common behaviour they are showing. Therefore, it is not a decision making action but rather more a routine-based reactive process. Nevertheless, knowledge of this typically Dutch behaviour, can help actors to plan their own, or joint activities (e.g., you can go to the supermarket at that time as it is likely to be very quiet, or you can invite your Dutch friends for dinner at that time). This is an example of a *shared strategy*, i.e. an institutional arrangement where different actors have the intention of performing the same task at a certain time or setting [17].

Even though the concept of shared strategy is socially and computationally very instrumental, it has not yet been implemented nor formalized in the MAS

J. Simão Sichman and H. Aldewereld (Eds.): COIN 2012, LNAI 7756, pp. 71–86, 2013.

literature. First, it determines the general behaviour of the system thus providing expectations about the global behaviour of the system. For example, restaurants should start preparing meals early since there will be many people coming at that time. Second, this notion adds a new dimension to the deontic classical concept where there is no obligation, permission or prohibition, yet a shared behaviour takes place.

In MAS research, shared strategies can be a new way of expressing conventions that cannot easily be fitted into norms as they have no deontic 'flavour' to it. Shared strategies are different from collective intentions [7]. A collective intention is a goal shared by everyone in a team. Moreover, members of the team are aware of other agents intention to meet the common goal. For a shared strategy however, while all agents possibly have the same goal, their execution of tasks to fulfil the goal are independent of each other and if one agent does not perform the task, their goal can still be met. For example, while two people may have the collective intention to watch a new movie together that has just been released, many people share the strategy to watch movies as soon as they are released. In the former case, the whole objective of watching the movie will not be achieved if the two do not manage to watch the movie together, in the latter however, whether one watches the movie does not effect the general goal.

Regarding the benefits of implementing the concept of shared strategies in MAS, in this paper we take inspiration from the Institutional Analysis and Development framework (IAD), an institutional economic framework developed by the Nobel laureate Elinor Ostrom [18]. IAD is an analysis framework for understanding social systems with the purpose of (re)designing social rules (i.e. norms). The ADICO structure, part of the IAD framework, provides a language for institutional statements, such as shared strategies, institutional rules and norms [6].

The remainder of the this paper is structured as follows. In section 2 we explore the different definition of institutions and introduce the IAD framework and ADICO statements. In section 3 we further define shared strategies and formalize the concept. Section 4 discusses how this definition can be used in MAS. Section 5 explores related work. Finally, section 6 gives our conclusions and directions for future research.

2 Institutions

Institutional economics is an area of research in the social sciences with a rich collection of theories and frameworks that can be highly instrumental for MAS research. Among these is the Institutional Analysis and Development (IAD) framework which has gained popularity in different disciplines. A major focus of this framework is institutional statements defined as the ADICO sequences. We take inspiration from the ADICO definition to formalize shared strategy for artificial agent societies.

In institutional economics, institutions are defined as "the set of rules actually used by a set of individuals to organize repetitive activities that produce outcomes affecting those individuals and potentially affecting others" [18,15]. These

rules include laws, regulations, social norms, and shared strategies amongst others. However, in MAS, the concept of institution usually refers not only to a set of rules as above but also to the regulative structures that enable verification and enforcement of those norms [4,1][1].

Institutions have two sides: on the one hand, they enable interactions, provide stability, certainty, and form the basis for trust. On the other hand, they may cause biased power relations. If institutions fail to fulfil stability or to enable decision making, there are grounds for institutional (re)design [13].

Institutional (re)design refers to the devising of new social arrangements, by examining existing arrangements and altering them when necessary [19]. I.e., institutional redesign refers to deliberate changes in institutional characteristics. In order to design institutions, one should be able to understand and analyse the institutional rules. Institutional frameworks such as the IAD framework by Ostrom [17] are developed for this purpose. This framework addresses the different components of a socio(-technical, -ecological) system that need to be analysed for institutional (re)design [17]. In the remainder of this section, we briefly introduce the IAD framework and the grammar of institutions (i.e. ADICO institutional statements).

2.1 Institutional Analysis and Design

The IAD decomposition of a social system is presented in figure 1. Its central concept is the 'action arena', in which individuals (or organizations) interact, exchange goods and services, engage in appropriation and provision activities, solve problems, or fight. The action arena is described by the participants (who have a set of resources, preferences, information, and selection criteria for action) and the action situation: the actual activity (or 'game') that is to be understood.

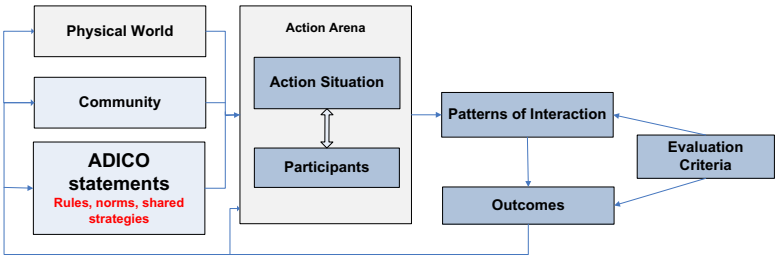

Fig. 1. The components of a social system in the IAD framework [18]

What happens in the action arena leads to patterns of interaction and outcomes that can be judged on the basis of evaluation criteria. The action arena itself is influenced by attributes of the physical world (e.g., climate, present technological artefacts), the attributes of the community in which the actors/actions

[1] Throughout this paper, we will be using the institutional economic terms where required.

are embedded (e.g. demographics), and the set of rules(referred to as institutional statements) that guide and govern the individuals behaviour.

Although physical world and community influence the action arena, it is the rules of the game or, in other words, the norms, that actually define it. Therefore, in IAD quite some attention is given to the institutional statements, which are decomposed into a structure (also referred to as grammar of institutions) called ADICO [17].

2.2 ADICO Institutional Statements

An ADICO statement consists of five components namely: Attributes, Deontic, aIm, Condition, and sanction (Or else). This decomposition is for the purpose of summarizing and analysing institutional statements[2], distinguishing between the different types and understanding the formation and evolution of these statements [17].

Attributes. Attributes describe the participants of an action situation to whom the institutional statement applies. Participants can be individuals who are distinguished by values such as age, sex or even roles in the system. For example, an attribute of an ADICO statement can be a 'student'. Corporate actors can also be considered as attributes instead of individuals (e.g., university). These actors can be distinguished by their organizational values such as location and size. The attribute component of an ADICO statement can never be empty. Therefore, if no attribute is specified for a given institutional statement the default value is 'all members' of the group.

Deontic Type. The purpose of the components is to distinguish between prescriptive and non-prescriptive statements. Deontic operators are *obligated* (O), *permitted* (P) and *forbidden* (H). While *obliged* and *forbidden* directly relate to the normative notions of 'ought', 'must' or 'should', *permitted* seems less related to the intuitive notion of norm. Permission rules however influence the structure of an action situation in three different ways. First by putting constraints on permissions and thus restricting actor behaviour. Second, assigning a permission to an action is constituting that action. Therefore, permission rules add action options to the action situation. Third, such rules grant rights to particular participants with certain properties to do an action. Some institutional statements don't have any deontic operator. As an example: "The person who places a phone call, calls back when the call gets disconnected".

Aim. The aim component describes the action or outcome (i.e., a state of affairs) to which the institutional statement applies. In order for a institutional statement to influence behavior, individuals must have a choice concerning its Aim. In other words, prescribing an action or outcome only makes sense if it its negation is also possible. E.g., the capability of voting implies the capability of not voting.

[2] We will use 'institutional statement' as a general term to address the concepts norm, rule and shared strategy.

Condition. Conditions are the set of parameters that define when and where an ADICO statement applies. If there is no condition stated it implies that the statement holds at all times.

Or Else. 'Or else' is the consequence of non-compliance to an assigned institutional statement. Only deontic statements include an 'Or Else'. A common type of 'Or else' is a sanction. Besides sanctions, rule violation may also result in the change of deontic (e.g. permitted to forbidden) of *another* rule. For example, it is forbidden to put a person in jail, but if they perform a crime, then the deontic changes to permission and one is allowed to imprison someone. Institutional actions may also be a result of norm violation. For example the role of the violator may be taken away. In general, the 'or else' component of an ADICO statement contains an institutional statement by itself which results in a nested structure of institutional statements. Also, the 'or else' component may be linked to the condition component that specifies the number of times that the norm has been violated.

According to the ADICO decomposition given above, an institutional statement can be divided into three different categories namely: rules, norms and shared strategies.

1. **ADICO**
   A *Rule*[3] (aka, regulatory rule) is the most complete form of statements covering all five components of the ADICO statement. In other words, rules have attributes, deontic type, action, condition and 'or else'.
2. **ADIC**
   A *Norm*[4] is an institutional statement without an 'or else' component. For example, shaking hands when being introduced to someone is a norm given that, if not done, it may affect your future relationship with that person. However, there is no fixed sanction and different people may have different reactions.
3. **AIC**
   A *Shared strategy* is an institutional statement where there are no sanctions or deontic type, and represents general expectations about the aggregate behaviour of others.

In the next section, we will discuss shared strategies in more detail.

3 Shared Strategies

3.1 Towards a Definition

According to E. Ostrom, a shared strategy is a social concept that refers to a type of behavioural pattern that is observed by a significant number of individuals although it is, prima facie, neither associated with any deontic modality,

[3] In agent literature, a rule is often addressed as 'norm' or 'regulation'.
[4] Sometimes called 'social norm' or even 'moral' or 'ethic code' in agent literature.

Table 1. Examples of Behaviours that can be assumed shared strategies

s_1	When a telephone conversation is cut, call back
s_2	When in Rome, do as Romans do
s_3	Dutch eat at 5:30
s_4	In a busy stairway, walk on the left
s_5	Jumping the queue is not nice
s_6	Faced with an unexpected obstacle, break
s_7	Only when a pedestrian makes a clear sign to attempt to cross the street, yield the right-of-way
s_8	If no police officer is in sight, skip the red light

nor having a reward or punishment linked to its performance. In order to elucidate the distinguishing features of shared strategies, in this section we explore different examples of social behaviour.[5]

Ostrom, in [17], pg. 143, proposes as an example of shared strategy, the rule of calling back when a telephone conversation is cut (s_1 in Table 1). Strategy s_1 is a conditional that under objective circumstances triggers an action. It does not explicitly entail an obligation or a prohibition, and no explicit or unique reward or punishment ensues. On a closer look, however, strategy s_1 may entail an *expectation*, that, depending on the context in which the interruption took place, may be a strong, possibly asymmetrical and, if not fulfilled may be consequential. The level and nature of expectation therefore reconciles with Ostrom's claim that, if an action rule is to be a shared strategy, then it would not matter whether α is done or not. We believe that the key is in the collective nature of expectations involved in shared strategies as we shall see.

Strategies s_2 and s_3 are similar to s_1 but their deontic component is more tenuous and thus closer to Ostrom's intuitive definition. Strategy s_2 "When in Rome, do as Romans do", like s_1, is an ostensible *directive for action* whose —relatively inconsequential— deontic component may guide the adaptive behaviour of foreigners, on one hand, and the leniency of natives towards non-standard behaviour of foreigners, on the other. Strategy s_3, "Dutch eat at 5:30", asserts a *factual regularity* but it also hides a directive for action whose compliance by an individual is indifferent to the rest of the world; nevertheless, under certain circumstances, it may have practical consequences (in Holland, for an individual's eating plans or for the operation of restaurants).

These three strategies may be deemed shared strategies only if we make some assumptions about the expectations involved explicit, otherwise they would be examples of *common* and *collective* strategies. Thus, strategy s_3 would be not a shared strategy but a "common strategy" if we understand it as a prevalent behaviour which people may not even be aware of. However, it becomes a "shared strategy" when we understand it as an expectation of common global behaviour;

[5] The concept of shared strategies has been addressed by social scientist using different terms. For instance, *scripts* by Schank and Abelson [21] or *conventions* by Hodgson and Knudsen [12]. For an overview of this literature see pg. 178 [17].

Table 2. Strategy Types

common strategy	most individuals do s_j
shared strategy	most individuals believe that most individuals do s_j
collective strategy	most individuals believe that most individuals believe most individuals do s_j

for instance, saying that most people believe that most Dutch eat at 5:30. Finally, s_1 also fails to be a shared strategy when the two parties expect that both parties should follow the rule, or technically, when there is collective belief. That is we have the three types of strategies characterized in Table 2:

Shared strategies may be situated, thus examples s_7 may only hold if in, say, Portugal. Furthermore, notice that some shared strategies (s_7 and s_8) may very well hold and be socially useful if situated in one context but may be highly dangerous patterns of behaviour in others, hence giving rise to full norms that forbid and punish their performance. Finally, situatedness is not only physical as s_5 "jumping the queue is not nice" illustrates. It is present in everyday situations like the supermarket or a theatre but becomes a strict directive (i.e., norm or rule) in surgery waiting lists and in some bureaucratic procedures.

As section 4 will show, it is important to distinguish between the collective character of a shared strategy –the fact that a collectivity has shared strategy or not— and whether each individual decides to enact or not that shared strategy in a particular moment. In fact, asymmetries of different types may create different expectations that affect agents' decisions; for instance, even when s_1 is a shared strategy, if I am calling a cab to go to the airport and communication breaks, it is me who should call back because it is in my best interest to continue the conversation and I may presume the cab doesn't know my number.

Likewise, shared strategies reveal a transient character that puts them between actual standard norms or social conventions, and fully unregulated behaviour, this transient character is revealed both in the collective and the individual perspectives. Thus, from an institutional perspective shared strategies can be seen as an *emerging social convention* or the *grounds for an emergent norm*. That is the case of s_4, "walking on the left of a busy stairway", that in London is a solid social convention —whose non-compliance is met with contempt or derision, while in Paris it is a shared strategy, and in the US it is not (still?). Note also that driving on one of either sides of a road, which was a shared strategy at some point, became institutionalized as an explicit norm everywhere; probably because of the social significance of non-compliance. From an individual's perspective, on the other hand, the transient character of shared strategies is evident in the same strategy s_4 that may be likened either to an *internalized norm* or to a *tacit social convention* of which the subject might be not fully aware.

3.2 Formalizing ADICO Statements

In this section we formalize the notion of *Institutional Statement* from Ostrom [17] to get to a semantic description of the rules, norms and, foremost, shared strategies of the ADICO framework. This forms the basis of our discussion on shared strategies in the next section.

The logic used for the formalization is a temporal epistemic logic based on CTL* [8] for the temporal aspect and KD45 [14] for the epistemic aspect. We use a technique similar to [9] for the combination of these modalities. In short, the resulting logic is a temporal logic where the states contain an epistemic modality. This allows for the expression of beliefs and changes of beliefs, but not the expression of beliefs about the temporal structures (that is, one can change its beliefs in a future state, but one cannot have beliefs about future or past states).

The core of the logic is given by the set of propositions $\mathcal{P}$, which can be used to construct sentences using the typical propositional operators $(\neg, \wedge, \vee, \rightarrow, \leftrightarrow)$. The set of all possible well-formed propositional formulas will be denoted as $\mathcal{L}_{\mathcal{P}}$. This logical core is extended to an epistemic logic of beliefs using a belief-operator (B), following the KD45 principles, resulting in a set of well-formed sentences $\mathcal{L}_{\mathcal{BP}}$. The temporal logical language $\mathcal{L}_{\mathcal{TBP}}$ is then constructed by adding the usual temporal operators: path operators A (all paths), E (some paths), and state operators X (next), G (always), F (sometime), U (until). The language is further enriched with *stit*: e_r ('see to it that', see [5]) to express individual action.

Using the logic $\mathcal{L}_{\mathcal{TBP}}$ we can now introduce the syntax of ADICO institutional statements as follows.

Definition 1 (Institutional Statement). *ADICO Institutional Statements are of the form*

$$D_R(I \,|\, C) \rightsquigarrow o$$

where
- D represents one of the modalities: $\{O, P, H, S\}$
- R being the attribute*, represented as a set of* roles*;*
- I being the aim*, represented as an expression from $\mathcal{L}_P$;*
- C being the condition*, represented as an expression from $\mathcal{L}_P$; and*
- $\rightsquigarrow o$ being the or-else*, where o is represented as combination of institutional statements.*

The modality of an institutional statement can either be: O (obligation), P (permission), H (prohibition), or S (shared strategy). The modality determines the semantics of the statement. Roles in our framework are considered as labels, with $\mathcal{R}$ being the set of all roles in the institution. The applicability of an institutional statement is thus $R \subseteq \mathcal{R}$. The $\rightsquigarrow o$ part of the statement expresses the *or-else* of the institutional statement, representing the reaction to violations of the statement. Intuitively, this means that when the lefthand-side of the $\rightsquigarrow$-operator is violated, the righthand-side of the $\rightsquigarrow$-operator is activated. The reaction, o, is represented as an expression containing institutional statements combined with

conjunctions and disjunctions. It is also possible that $o \equiv \top$, which expresses that the institutional statement has no reaction[6].

The different types of institutional statements referred to by Ostrom can be obtained in the following ways. A *rule* is an institutional statement that contains all elements, and where the modality is of deontic nature (that is, $D \in \{O, P, H\}$). *Norms* are institutional statements with a deontic modality ($D \in \{O, P, H\}$) and where no o is specified; $D_R(I \,|\, C)$. Finally, *shared strategies* are institutional statements without a deontic modality ($D = S$) and where the reaction o is absent; $S_R(I \,|\, C)$.

For the semantics of the institutional statements, we create reductions of the newly introduced operators to the basics of the $\mathcal{L}_{\mathcal{TBP}}$. Due to space limitations, we give the reduction of obligations, prohibitions and shared strategies; the reduction of permissions (weak permissions, strong permissions, cf. [23]) is out of scope of this paper, and left as an exercise to the reader.

Definition 2 (Reduction of Obligations).

$$\begin{aligned} O_R(I \,|\, C) \rightsquigarrow o \Leftrightarrow \forall r \in R A\Big[& C \rightarrow (\neg viol(I, r))\ U \\ & \big(e_r\, I \wedge X(AF \neg viol(I, r)) \vee \\ & X(\neg I \wedge viol(I, r))\big) \\ & \wedge viol(I, r) \rightarrow o\Big] \end{aligned}$$

The above definition transforms the obligation into a $\mathcal{L}_{\mathcal{TBP}}$ sentence, using an Anderson's reduction [2], similarly as done in, e.g., [1]. Intuitively, the definition expresses that whenever the condition (C) holds, *either* the aim (I) is achieved by those obliged ($e_r\, I$), in which case no violation of the obligation will ever occur, *or* the aim is not achieved, and a violation happens. Moreover, when the violation happens, the reaction statement o (if present) is triggered (these statements typically express sanctioning mechanisms, see [17]).

Definition 3 (Reduction of Prohibitions)

$$H_R(I \,|\, C) \rightsquigarrow o \Leftrightarrow O_R(\neg I \,|\, C) \rightsquigarrow o$$

The reduction of prohibitions is based on the principle that $Hp \equiv O\neg p$ from most deontic logics.

Definition 4 (Reduction of Shared Strategy)

$$S_R(I \,|\, C) \Leftrightarrow \forall r_1 \in R, \forall r_2 \in R \backslash \{r_1\} : A(C \rightarrow B_{r_1} e_{r_2} I)$$

The reduction of shared strategies is formed around the idea that shared strategies represent an *expectation*. Intuitively, a shared strategy expresses the expectation that other members of the same group (i.e., playing the same role, or

[6] Typically, when $o \equiv \top$, we omit the $\rightsquigarrow o$ part of an institutional statement for readability: $D_R(I \,|\, C) \rightsquigarrow \top = D_R(I \,|\, C)$.

part of the group of roles that share the strategy) will try to follow the shared strategy. This idea is reflected in definition 4. This is different from the notions of common strategy, where everyone in the group does the expected thing, and joint strategies, where everyone in the group intends that they do the expected thing. Using similar elements as used in definition 4, we can also formalize the notions of common strategy and joint strategy:

Proposition 1 (Common & Joint Strategies)

$$CS_R(I \mid C) \Leftrightarrow A(C \rightarrow \forall r_1 \in R : e_{r_1} I)$$
$$JS_R(I \mid C) \Leftrightarrow A(C \rightarrow \forall r_1, r_2 \in R : B_{r_1} B_{r_2} e_{r_1} I)$$

Common strategies (CS) happen when all agents in a system are programmed alike, and act in similar manners; that is, every member of a group R follows a common strategy CS_R to do I when each member of that group does I. A joint strategy (JS), similar to joint-intentions [7], is when every member of a group R does I, but also knows (and expects) that every other member of R also does I. That is, there is shared belief that the group beliefs that they are doing I.

By formalizing the shared strategies (and similarly, common and joint strategies) we lost an aspect of Ostrom's concept. An important aspect of Ostrom's reading is that a shared strategy can be not acted upon, which is missing from definition 4, since we expect that every agent in the group will do I. Informally, definition 4 reads as "everyone from group R believes everyone from group R does I". Ostrom's reading of a shared strategy is more in line with "most from group R believe that most of group R do I" (see the discussion earlier in section 3.1). This has an impact on the way agents behave, because in the first reading one can be sure that members of the group R will do I, whereas in the second reading it might be that some members of R will not do I. Therefore, we need to weaken our definition, for which we require a semantic definition of 'most'.

Definition 5 (Most). *We define the set-theoretic 'most' operator* W *as follows, for a set of roles* R*:*

$$\mathsf{W}(R) = R' \Leftrightarrow R' \subseteq R \wedge (|R'| > 1/2 \cdot |R|)$$

Intuitively, this definition expresses what one would expect. If R' is representing the most of set R, then at least half of the agents in R are also in R'; that is, R' is a subset of R and the number of elements of R' is at least half that of R.

Using the concept of 'most' we can create weaker versions of the earlier strategies as follows.

Proposition 2 (Weak strategies)

$$CS^-_R(I \mid C) \Leftrightarrow A(C \rightarrow \forall r_1 \in \mathsf{W}(R) : e_{r_1} I)$$
$$JS^-_R(I \mid C) \Leftrightarrow A(C \rightarrow \forall r_1, r_2 \in \mathsf{W}(R) : B_{r_1} B_{r_2} e_{r_1} I)$$
$$S^-_R(I \mid C) \Leftrightarrow$$
$$A(C \rightarrow \forall r_1 \in \mathsf{W}(R), \forall r_2 \in \mathsf{W}(R \backslash \{r_1\}) : B_{r_1} e_{r_2} I)$$

Table 3. Examples of Shared Strategies

s_1	$S^{-}_{on\_phone}(call\_back \mid conversation\_cut)$
s_2	$S^{-}_{tourist}(do\_as\_Roman \mid in\_Rome)$
s_3	$S^{-}_{Dutch}(eat \mid 5:30)$
s_4	$S^{-}_{pedestrian}(stay\_left \mid in\_busy\_stairway)$
s_5	$S^{-}_{civilised\_people}(\neg jump\_queue)$
s_8	$S^{-}_{driver}(skip\_red\_light \mid \neg police\_in\_sight)$

The expressions in proposition 2 represent the weakened versions of the expressions in proposition 1 and definition 4. Intuitively, they read as follows. A group R has a *weak common strategy* to I when most of R do I. A group R has a *weak joint strategy* to I when most members of R believe that most other members of R believe that most of them do I. Finally, a group R has a *weak shared strategy* to I when most members of group R believe that most other members of R do I.

A formalization of some of the examples from table 1 is shown in table 3 below.

4 Shared Strategies Applied in MAS

In this section, we discuss the practical application of shared strategies in MAS. Shared strategies can be seen as a form of regulation of individual behaviour within a system, or as mechanisms to improve cooperation, coordination and control in MAS. As such, shared strategies can be used by agents in their reasoning processes, in order to determine their plans in a shared environment (cf. section 4.1), or as means to support design and evaluation of engineered MAS (cf. section 4.2).

4.1 Individual Application

In this section, we look at how shared strategies can be used by individual agents in their planning. As with norms, agents can and should take into account the shared strategies holding in a domain in order to generate efficient plans for their goals. We assume here autonomous cognitive agents that are able to use their knowledge about a domain in the generation of plans. Such agents can decide on the adherence or not to norms. Other researchers have studied norm-based planning [22], i.e. the generation of optimal plans with respect to a set of norms. In this section, we concentrate on the use of shared strategies for the generation of plans.

The intuition of the formal definition of shared strategy introduced in section 3 is that most agents assume that under certain conditions, other agents will behave in a certain way. While common strategies may be designed into agent systems so that agents are not aware of them as common behaviour, shared strategies can be perceived by the agents as shared behavioural patterns. If most agents see that most

agents have this new perception, the strategies will be globally recognized as shared strategies. This new knowledge will then be updated in their belief system and used in their planning. Based on these new beliefs, agents can take two approaches to use shared strategies in their planning, referred here as an *optimistic* and a *pessimistic* approach. In order to discuss the difference between these two approaches, we take as example the shared strategy:

$$S^{-}_{drivers}(break|obstacle\_in\_road)$$

which represents the fact that drivers will break when there is an obstacle in the road.

An optimistic pedestrian agent will assume that all drivers will break when she crosses the road, and therefore will plan to cross the road even if she sees a car approaching. On the other hand, a pessimistic pedestrian will assume that you cannot know which drivers will adhere to the shared strategy, since not all have to follow it, and therefore will plan to stop at the curb when she sees a car approaching.

We are currently working on an extension to the BDI architecture that incorporates reasoning using shared strategies.

4.2 Institutional Application

Form an institutional perspective there are two issues worth identifying. The relationships between shared strategies and institutional design and evolution, and the role of shared strategies in multiagent-based simulation.

Since shared strategies constitute a regularity of the aggregate behavior, institutional conventions may be designed to promote or to control the consequences of that regularity. The approach is straightforward when the existence of a shared strategy is known in advance and it is likely that its execution carries out institutional objectives. In this case, it is reasonable to include specific evaluation mechanisms to monitor the effects of the strategy, and use these to assess transaction costs that would in turn guide the adaptation of the institution to actual performance ([14]). Concomitantly, it is also feasible to establish institutional norms and conventions —with the appropriate evaluation mechanisms— that regiment, constrain or foster the enactment of the shared strategy by participating agents.

The way of dealing with the alternative case is less obvious. When the existence of a shared strategy is not known in advance, ordinary performance monitoring does not necessarily identify the behavioral regularity, even when performance indicators might signal a hidden cost. In such case, institutional reaction may be untimely and inefectual. To contend with such eventuality, one may attempt to foresee undesirable outcomes and, at the risk of overregulation, legislate against them. The opacity of undesirable outcomes, however, may sometimes be appropriately addressed with conventional mechanism-design techniques or by a clever use of modeling and simulation methodologies.

In addition to their value for visualizing the effect of shared strategies on institutional performance, in this context, the modeler deals with the system as

a regulated MAS, making a shared strategy a feature of individual agents and harnessing individual actions through institutional conventions of different sorts. The use of shared strategies may be fruitful for some forms of agent-based simulation. One relevant form is to use shared strategies as a salient part of the agents' internal decision models. This way, the designer may study different aspects of normative, motivational and goal-directed attitudes (for example the interplay of norms and strategies in different agent architectures, norm internalization processes, norm emergence, norm compliance vs. conflict resolution approaches, value formation, achievement degrees). Another form of using shared strategies in agent-based simulation is to factor the analysis of aggregate behavior by designing populations partitioned by shared strategies, thus measuring cost and value of interactions within populations with pure and mixed strategies, rational or spontaneous triggering of the shared strategies, etc.

4.3 Institutional Emergence

Although this is not the primary focus of this paper, we see the ADICO structure as an instrumental tool to study the emergence of rules, norms and shared strategy in agent societies.

As Ostrom explains in [17], the change in any part of the ADICO statement results in the evolution of such entities in a society. For example, when global expectations about a shared strategy narrow down to individuals, a deontic flavour emerges, turning the shared strategy to a norm. Likewise, when the implicit, non-unique and unclear consequences of non-compliance to a norm become common, known and explicit to everyone, that norm turns into a sanction.

Besides the study of institutional evolution as we described above, the ADICO statement can be linked to the internal architecture of an agent (e.g. BDI) so that the agents can perceive common behaviour and recognize and establish it as an institutional statement. For example, if the agent detects the components of an ADICO structure in a repeated pattern of behaviour in the society it will announce this as a shared strategy/rule/norm, and if many agents announce the same statement, this will become an emergent ADICO statement in the artificial society. We are only addressing the idea of this application. However, the implementation will be the topic for future work.

5 Related Work

Some concepts in the MAS literature address shared strategy to some extent. Table 4 shows some of the most relevant concepts and compares their usage with similar examples. Normative information can be situated in the environment (e.g. sign boards) which means that a norm only needs to be followed within a certain boundary of space and time [16]. The type of situated norm can be warning, obligation and direction. A shared strategy however, does not necessarily have to be bound to location and time or have any of the types given to distributed norms (i.e. warning, obligation, direction).

Social conventions are rules that restrict agent behavior while having no threat or punishment. Young (1993) presents the following definition of a conventional norm: "A convention is a pattern of behavior that is customary, expected, and self-enforcing. Everyone conforms, everyone expects others to conform, and everyone wants to conform given that everyone else conforms."

For a shared strategy however, no one has expectation for others to conform because they are not aware if the person is necessarily a follower of the strategy. No (low) expectation results in no (low) disappointment. For example, if in a given context calling back if the line is dropped is a social convention, then the person may be upset but if it is a shared strategy, the person does not know if he is a performer of the shared strategy 'calling back', and thus will not be offended if the caller does not call back. Therefore it can also be concluded that a shared strategy has lower priority than a convention for agent planning.

A collective intention is the reason for team existence and it implies that all members intend for *all* others to follow that intention [7] . The goal of the team may not be reached if one agent may decide not to follow the intention. However, for a shared strategy, as mentioned previously, most people know the strategy and know that most others will follow the strategy. Therefore, there is no obligation for agents to perform the strategy and there is also no significant consequence on an individual level while the global behavior of the system may be important.

Table 4. Concepts related to shared strategy in current MAS literature

Concept name	**ref.**	**Example**
Shared Strategy	[17]	The Dutch eat dinner at 5:30pm.
Situated Norm	[16]	In this ship dinner is served at 5:30 pm (or else no food).
Social Norms/ Conventions	[24]	When eating dinner, people start at the same time
Shared/Collaborative plans	[11,10]	Those group of friend have plan to make dinner together
Collective Intention	[7]	Those group of friend are committed to have dinner together at 5:30 pm.

Norm internalization [3] is another topic of research in MAS that can be used in combination with shared strategies. Norm internalization is progressive. This is in line with the transition of ADICO statement from one type to another (e.g. a norm becomes a shared strategy)[16, 3]. In other words, during the process of internalization, an ADICO rule which has all five parts of the statement, may loose the 'or else' and become a norm and later on turn into a 'shared strategy' by losing the deontic. On the other hand, the more the norm is internalized the less decision making is required. This again is in line with the definition of shared strategy which is more of a routine that requires less thinking. A fully internalized norm is a shared strategy only if it is shared among people.

The original formulation of shared plans [10] does not see the necessity for an agent to have intentions towards the act of another agent. It is similar to shared strategies in the sense that there is not joint intention between the agents. However, it is different to shared strategies because the agents make plans and actually coordinate in performing the action. Collaborative plans [11] which are a revised version of shared plans are also different from shared strategies because they produce commitment to the joint activity.

6 Conclusion and Future Work

In this paper we presented the concept of shared strategy as an alternative concept to that of norm in MAS. Based on the work of Ostrom, namely the notion of ADICO institutional statement, we presented an integrated formalism to describe the semantics of norms and shared strategies, based on a temporal epistemic logic.

A shared strategy is a low priority statement leading to action among a group of agents. Since the expectation is *shared*, each agent believes that *most* other agents will perform the action but does not necessarily know who. Therefore, agents don't have expectations for a particular other agent to perform shared strategies because they cannot know whether that particular agent follows the strategy or not, even though as a group, most will. This yields that no deontic type and no sanction can be assigned to a shared strategy.

Shared strategies are a crucial part of agent societies as they result in global behaviors that may need to be taken into consideration by other agents who may be part of the system or merely global viewers. A shared strategy can change into norm and vice versa depending on the level of norm internalization and the context which facilitates the implementation of norm emergence and evolution [20].

For future work, we are further extending the formalization of shared strategy. We are also exploring how shared strategies can be implemented into BDI architecture.

Acknowledgments. This research was partly funded by the COST Action IC0801 "Agreement Technologies" Grant.

References

1. Aldewereld, H.: Autonomy vs. Conformity: an Institutional Perspective on Norms and Protocols. SIKS Dissertation Series 2007-10. Utrecht University, PhD Thesis (2007)
2. Anderson, A.: A reduction of deontic logic to alethic modal logic. Mind 67, 100–103 (1958)
3. Andrighetto, G., Villatoro, D., Conte, R.: Norm internalization in artificial societies. AI Communications 23(4), 325–339 (2010)
4. Arcos, J., Esteva, M., Noriega, P., Rodrguez, J., Sierra, C.: Engineering open environments with electronic institutions. Journal on Engineering Applications of Artificial Intelligence 18(2), 191–204 (2005)

5. Belnap, N., Perloff, M.: Seeing to it that: a canonical form for agentives. Theoria 54(3), 175–199 (1988)
6. Crawford, S., Ostrom, E.: A grammar of institutions. American Political Science Review, 582–600 (1995)
7. Dunin-Keplicz, B., Verbrugge, R.: Collective intentions. Fundamenta Informaticae 51(3), 271–295 (2002)
8. Emerson, E.: Temporal and modal logic. In: van Leeuwen, J. (ed.) Handbook of Theoretical Computer Science, vol. B, pp. 955–1072. MIT Press (1990)
9. Engelfriet, J.: Minimal temporal epistemic logic. Notre Dame Journal of Formal Logic 37(2), 233–259 (1996)
10. Grosz, B., Kraus, S.: Collaborative plans for complex group action. Artificial Intelligence 86(2), 269–357 (1996)
11. Grosz, B., Sidner, C.: Plans for discourse. Technical report, DTIC Document (1988)
12. Hodgson, G., Knudsen, T.: The complex evolution of a simple traffic convention: the functions and implications of habit. Journal of Economic Behavior & Organization 54(1), 19–47 (2004)
13. Klijn, E., Koppenjan, J.: Institutional design. Public Management Review 8(1), 141–160 (2006)
14. Meyer, J.-J.C., van der Hoek, W.: Epistemic Logic for AI and Computer Science. Cambridge University Press (1995)
15. North, D.: Institutions, institutional change and economic performance. Cambridge University Press (2009)
16. Okuyama, F., Bordini, R., da Rocha Costa, A.: Spatially distributed normative objects. In: Coordination, Organizations, Institutions, and Norms in Agent Systems II, pp. 133–146 (2007)
17. Ostrom, E.: Understanding institutional diversity. Princeton Univ. Pr. (2005)
18. Ostrom, E., Gardner, R., Walker, J.: Rules, games, and common-pool resources. Univ. of Michigan Pr. (1994)
19. Pettit, P.: Institutional Design and Rational Choice, pp. 54–89. Cambridge University Press (1996)
20. Savarimuthu, B., Cranefield, S., Purvis, M., Purvis, M.: Norm emergence in agent societies formed by dynamically changing networks. Web Intelligence and Agent Systems 7(3), 223–232 (2009)
21. Schank, R., Abelson, R., et al.: Scripts, plans, goals and understanding: An inquiry into human knowledge structures, vol. 2. Lawrence Erlbaum Associates, Nueva Jersey (1977)
22. Panagiotidi, S., Vázquez-Salceda, J.: Normative Planning: Semantics and Implementation. In: 13th International Workshop on Coordination, Organizations, Institutions and Norms in Agent Systems (COIN@WI-IAT), Lyon, France (2011)
23. van der Torre, L.: Deontic Redundancy: A Fundamental Challenge for Deontic Logic. In: Governatori, G., Sartor, G. (eds.) DEON 2010. LNCS, vol. 6181, pp. 11–32. Springer, Heidelberg (2010)
24. Villatoro, D., Sen, S., Sabater-Mir, J.: Of social norms and sanctioning: A game theoretical overview. International Journal of Agent Technologies and Systems (IJATS) 2(1), 1–15 (2010)