

From roles to standards: a dynamic maintenance approach using incentives

Ramón Hermoso · Henrique Lopes Cardoso ·
Maria Fasli

© Springer Science+Business Media New York 2014

Abstract Social coordination has been addressed in multi-agent systems, making use of concepts such as institutions, norms, commitments, conventions, roles, or trust. In this paper, we argue the need to tackle open and dynamic environments with yet another concept: the notion of a standard, seen as a measurable and non-committing expectation. Not much work has been done in the field of multi-agent systems addressing the evolving nature of roles, especially in open systems, in which changes in the population bring about changes in the expectations generated from roles. Using standards measured from roles as the focus of attention, we propose an incentive-based mechanism to maintain roles over time. This approach is put in contrast with reorganization, which is needed when incentives are not cost-effective. Different search algorithms are proposed to illustrate incentive-based maintenance. Some empirical results are shown based on the principal-agent model from economics.

Keywords Artificial societies · Standards · Incentives

R. Hermoso (✉) · M. Fasli
School of Computer Science and Electronic Engineering,
University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK
e-mail: rhermoso@essex.ac.uk

M. Fasli
e-mail: mfasli@essex.ac.uk

H. Lopes Cardoso
LIACC / DEI, Faculdade de Engenharia, Universidade do Porto,
Rua Dr. Roberto Frias, 4200-465 Porto, Portugal
e-mail: hlc@fe.up.pt

1 Introduction

Since the mid 1990s, a considerable number of works have been concerned with the development of infrastructures for supporting social coordination in open multi-agent systems. Taking inspiration from social sciences, concepts such as (electronic) institutions (Esteva et al. 2001; Dignum and Dignum 2001; Fornara et al. 2008), norms (Boman 1999; López y López and Luck 2003; Boella et al. 2006; García-Camino et al. 2007; Lopes Cardoso and Oliveira 2008), commitments (Singh 1999; Fornara and Colombetti 2003; Fasli 2003), conventions (Walker and Wooldridge 1995; Conte and Castelfranchi 1999), and roles (Hubner et al. 2002; Fasli 2006; Winikoff and Craneffeld 2009; Hermoso et al. 2013) have been introduced and exploited in the multi-agent systems domain.

In open multi-agent systems, agents enter and leave the interaction environment, and behave in an autonomous and not necessarily cooperative manner, exhibiting self-interested behaviours. Even when agents establish commitments among them, the dynamic nature of the environment may jeopardize such commitments if agents are not socially concerned enough and value more their private goals when evaluating the new circumstances.

Moreover, in open and dynamic environments one cannot assume that agents will behave consistently over time. This may happen either because of the agents' (lack of) ability or benevolence attitude. In some cases, an agent may not be capable of maintaining a certain behaviour standard throughout its lifetime. In other cases, the agent may intentionally deviate from its previous performance. It is therefore important, when considering open environments, to take into account also the evolution of an agent's internal skills or motivations, besides the dynamics of the interaction environment as a whole.

These kind of issues have been the motivation for the development of computational trust models, which may be used to enable an informed selection of an interaction peer. Some of these approaches comprise *situational* trust models (Rehak et al. 2006; Tavakolifard et al. 2008; Hermoso et al. 2013; Urbano et al. 2011), in the sense that agents are evaluated regarding their performance in specific contexts, situations, or tasks. A typical assumption in these models is that the measured trustworthiness of an agent is updated as new experiences and evidence are collected from the environment. There is therefore no collective perspective on the group of evaluated agents – each is assessed individually within the multi-agent system, and therefore no overall evaluation is performed.

Taking an organizational approach, and looking at the society from a role-specialization perspective, Hermoso et al. proposed role evolution (Hermoso et al. 2013) as a guideline to develop a coordination mechanism that enhances partner selection processes for task delegation purposes. This approach is based on examining the agent society and on identifying “run-time roles” – so building a role taxonomy – that clusters agents with similar skill patterns for a certain (set of) task(s). From this perspective, the mechanism provides, as a service, the identification of the role that labels agents considered as the most suitable to perform a specific task.

Looking at this role taxonomy as providing structure for some sort of an artificial organization, in this paper we address the problem of organizational maintenance. Given the evolving nature of agents, as pointed above, a problem faced by the organization in which agents have been (artificially) embedded is that of timeliness: are the agents within a role still performing as well as they did at the time of their assessment? One of two possible options can be chosen when agents start under-performing. The first is to reorganize so that the role taxonomy becomes accurate again. But assuming that this reorganization may be costly, the second approach is to influence the agents’ reasoning by making use of incentives or punishments, in an attempt to keep them on track. We propose and exploit the concept of *standard* to capture the level of suitable performance that agents have to show when they carry out different tasks.

The rest of the paper is structured as follows. Section 2 provides an overview of related work, looking at several means of addressing the problem of social coordination, and providing support for the concept of standard. Section 3 summarises the work on role evolution this paper is based on and presents a “standardisation” process from roles to create standards as objectively measured performances to be maintained for each role. In Section 4, we describe the action apparatus and decision rationale of agents based on the principal-agent model. Then, in Section 5, we put forward a model to establish and adjust incentives in order to

maintain standards over time. Different search algorithms are proposed in this regard. Section 6 discusses the need for reorganization and describes a rationale and an approach to undertake it. We evaluate our proposals and present empirical results in Section 7. Section 8 provides a critical discussion of the contributions of this paper and compares it with other related approaches in the literature. Finally, we conclude and sketch our planned future work in Section 9.

2 Related work

This paper proposes an approach to tackle open and dynamic environments, based on the notions of role and standard. Social coordination is a very active topic in the multi-agent systems research community. The need for approaches to tackle the open and dynamic nature of multi-agent environments has given rise to complementary approaches that bring into this realm concepts from diverse fields, such as social, organizational, legal, or behavioural sciences.

The concept of *institution* has been borrowed from economics (North 1990) and philosophy (Searle 1995) in two directions. On the one hand, to provide regulated environments consisting of computational infrastructures that frame agent interactions (as e.g. in Esteva et al. (2001) and Lopes Cardoso and Oliveira (2008)). On the other, to provide semantics to agent interactions in terms of counts-as relations (Jones and Sergot 1996; Fornara et al. 2008).

Within these regulated environments the notion of *norms* (Boella et al. 2006) has been exploited, as a means to explicitly state what to expect from each agent in the system. Social norms (Tuomela 1995) are based on mutual belief, consisting of *conventions* (Walker and Wooldridge 1995; Conte and Castelfranchi 1999) that may apply to a large group of agents.

The specification of *roles* (Hubner et al. 2002; Fasli 2006; Hermoso et al. 2010) that agents enact in a given society goes in the same direction of making a system more predictable in terms of expected behaviour (Winikoff and Cranefield 2009). Explicitly handling such expectations as norms (Castelfranchi et al. 2003) allows one to make role enacting agents accountable for their actions.

Although closely associated with some approaches on the use of norms, the notion of *commitment* (Singh 1999; Fornara and Colombetti 2003; Fasli 2003) emphasizes a deliberative view on norm adoption: agents commit to this as a result of their deliberation process, as opposed to a top-down view on the use of norms as a design tool for defining rules of behaviour in an interaction environment.

What seems to be missing from these artifacts for social coordination is some concept of expectation that is both measurable and non-committing. Unlike commitments or

norms, we pursue a means of establishing how well an agent is able to perform without actually being committed to. And unlike conventions, which typically apply to collective behaviour, we are interested in measuring the outcome of task executions. In order to fill this gap, we use the concept of *standard*.

In the literature, one can find a wide variety of works that adopt the concept of standard for different purposes. The definition of *standard* is given by the Oxford Dictionary of English¹ as: “(noun) 1. a level of quality or attainment; 2. something used as a measure, norm, or model in comparative evaluations; (adj.) 3. used or accepted as normal or average”. Therefore, standards describe levels of quality that are recognized to be normal by individuals within a type of system.

Standards come about either imposed or emerge as *de facto* after a number of observations. Following this distinction, standards may be classified as *de jure* or *de facto* standards (Salgé 2005). The former are regulations accepted and obliged by law, and are endorsed by a formal standards organization. An example of this type of standards are IEEE or ISO standards for many different purposes. In contrast, *de facto* standards arise when a critical mass simply agrees to use them in a particular environment. For example, PDF became a *de facto* standard for printable web documents, although it turned into a *de jure* standard as ISO 19005-1 in 2005. In this paper, we work on *de facto* standards in order to capture the measured performance of a set of individuals.

In the field of economics, we can find many approaches dealing with standards. In Busch (2000), Busch suggests that standards are mistakenly considered to be mere convenient technologies for organizing and regulating markets so as to reduce transaction costs (however the cost is assessed). Nevertheless, the author argues that standards are part of the moral economy of a society, since they also regulate behaviour. Therefore, standards create uniformity in heterogeneous contexts. Following up the economic approach, Murphy claims that performance standards emerge from the desire to provide incentives while simultaneously paying competitive expected levels of compensation (Murphy 2000). That is, performance standards can be used in order to gauge the adequacy of an individual's performance, so prompting the possibility of offering incentives if behavioural deviation exists. From a psychological point of view, Bandura and Wood (Bandura and Wood 1989) claim that when people believe the environment is controllable on matters of importance to them, they are motivated to perform as well as they can, which entails an increase in the likelihood of success (or societal satisfaction). Moreover, successful experiences, in turn, provide a self-validation

on the efficacy of the individual and of the environmental controllability. On the contrary, if people face situations they believe as being uncontrollable, they are likely to put in less effort, and this brings about failure. Consequently, over time, failures take an increasing toll on perceived self-efficacy and beliefs about how much environmental control is possible (Bandura and Wood 1989). This author claims that rates of success and failure are largely determined by the standards against which attainments are gauged. Therefore, in layman's terms, performance standards are used in order to induce collaborative behaviour from counterparts engaged in an interaction. However, as we state in this paper, performance standards *per se* are not sufficient to keep individuals from performing inadequately. This feature drives us to introduce the concept of incentive as a means to maintaining the performance standards. How standards are created has been studied by many researchers from different fields as well. This process is called *standard setting* and concerns the methods to build standards from the information available (or potentially available) in the systems (Cizek and Bunch 2007). There exists a wide range of approaches with regards to standard setting, from educational purposes (Hambleton et al. 2000) to medical evaluations (Southgate et al. 2001).

3 Creating standards from roles

Due to the non-stationary nature of open systems, in this paper, we address the problem of how to maintain agents from a performance quality perspective as determined by the roles they are playing. We claim that the notion of specialized role proposed in Hermoso et al. (2013) might be used to establish performance standards and so facilitate the agreement on commitments among agents. We therefore envision the possibility of going from roles as expectations of behaviour (Winikoff and Cranefield 2009) or performance (Hermoso et al. 2013) to the explicit handling of such expectations as *de facto* standards that may be committed to. Standards are therefore not imposed by the system, but promoted after they have been identified as current practice among a group of agents. Standards may thus be exploited to find appropriate agents, given that they are based on evidence about their actual capabilities.

The rationale behind creating and maintaining performance standards relies on the concept of role proposed by Hermoso et al. (2013). In this work, the authors claim that in open systems the evolution of the population should entail that organisational structures evolve as well. For instance, some of the roles that existed in our society two centuries ago, no longer exist nowadays. Therefore, roles are somehow linked to the different needs of the population at a certain point in time. With this idea in mind, the authors

¹<http://oxforddictionaries.com/>

define roles as entities that group a set of agents that outperform others for a certain set of tasks. Furthermore, they introduce the concept of specialisation, from which roles are created as specialisations of more general roles; e.g. the role surgeon is created as a specialisation of the role doctor, because those playing the former (they also play the role doctor) are better skilled for tasks such as *operate*. Thus, the authors propose that any society of agents may be covered by an overlay role taxonomy formed by extracting capabilities and trust relationships among agents over time.

The mechanism evolves the society's role taxonomy, assigning agents to roles.

Then, roles other that agents are playing in the system provide information about their expected capabilities regarding certain interactions (e.g. the provisioning of certain services or tasks). The authors assume that agents participating in the system are rational, that is, they behave as utility maximisers. Thus, the main task of the mechanism is twofold: i) to capture similar behaviour among participants that play a role; and ii) to manage the role taxonomy that structures different positions of agents in the system. The mechanism uses a clustering algorithm to identify patterns of behaviour, so distinguishing those agents outperforming others.⁰ This mechanism has been exhaustively tested in different conditions with open task-oriented multi-agent systems with heterogeneous and dynamic populations, showing a significantly good adaptation in order to provide an efficient role taxonomy that improves agents' partner selection.

The mechanism in Hermoso et al. (2013) relies on the definition of a Task-oriented Multi-Agent System (T-MAS) as a multi-agent system in which participants have to perform a set of tasks (Hermoso et al. 2013). It is defined by a set of participants \mathcal{Ag} , a set of tasks \mathcal{T} and a role taxonomy Δ . We will use the notion of T-MAS throughout the paper as the base for our approach.

The role taxonomy reflects, at a given time, which agents are more skilled in the system to perform different tasks. As the process of role taxonomy evolution is costly, in this paper, we focus on the period among two consecutive evolutions. Given a role taxonomy, we will extract performance standards from the roles in it. These standards will emerge as an indicator of what is expected from the group of agents (the ones playing a specific role) for a particular task. Thus, standards will be used by agents as an anticipatory measure on the likely outcome of interactions, and may be used in further negotiations to assure a certain quality of performance.

Standard setting process Let $Attr = \{attr_1, attr_2, \dots, attr_n\}$ be the set of attributes that characterizes a task in \mathcal{T} . For instance, in an e-commerce domain $Attr = \{delivery_time, quality_type\}$ would be a set of attributes that

might characterize the task *supply_good*. Let x_i be a value for the attribute $attr_i$. For example, $x_{delivery_time} = 5$ means that the value for the attribute *delivery_time* is 5. Thus, let us define the concept of standard:

Definition 1 A **standard** ς is a tuple $\langle r, t, x_i \rangle$ that establishes, for a role r and a task t , a certain expected level of quality x_i for the attribute $attr_i$ in t .

Note that the concept of standard refers only to one attribute of the task. That is, the same role and the same tasks might have different standards for different attributes. Following up the example given above, the standard for *delivery_time* might be 5 while the standard for *quality_type* might be 3.

The level of quality of different attributes is meant to be a target (x_i) representing the expectation an agent performing a task generates on its counterparts. In order to avoid our approach being domain-dependent, we take this notion of targeted standard to be as abstract as possible; that is, we cannot state a priori in which situations (in any type of domain), higher or lower outcomes (regarding the standard value) bring about better or worse outcomes. As we will further explain in Section 4.1, a behaviour showing a deviation exceeding the standard is equally harmful than the one that does not reach it. Thus if the standard for *delivery_time* is 5 days and a provider takes 7 days delivering a product, that is considered as undesirable as providing the product in 3 days. Although this would seem to be a bit counter-intuitive, delivery in advance might mean no storage room for the new products, while late delivery might entail that the production line will be halted.

In this paper, we claim that once the role evolution mechanism described above is in place, roles may be used to create standards which, in turn, could be included in commitments that regulate interactions in the system. The underlying idea relies on the aggregation of values of different attributes that characterize (the performance of) a task within the role, to establish a standard for that role/task pair, as defined in definition 1. For instance, in open electronic markets (e.g. eBay), roles might be created in order to place providers in different categories of provision, while standards would emerge in order to facilitate better interaction processes, and so prevent agents from exhibiting undesirable behaviours, such as longer delivery times, price changes, decrease of quality, etc. More precisely, let us suppose we have the role *Bike Provider* with, among others, the attributes *delivery time* and *quality_type*. Let a_1 , a_2 and a_3 be three agents playing the role, with average delivery times of 3, 4 and 5, respectively. Then we could use an aggregation function (e.g. an average) to extract a standard for the task as $\varsigma = \langle r = \textit{Bike Provider}, t = \textit{Provide Bikes}, x_{delivery_time} = 4 \rangle$. Using an averaging function may make

sense if we take into account that the clustering mechanism obtaining the role taxonomy will, in principle, get us roles with high cohesion – for which not much deviation should be expected in the beginning.

4 Incentives and the principal-agent model

After explaining the path from roles as expectations to commitments based on standards, we are now in a position to elaborate on enforcement schemes that enable us to maintain the stability of the role taxonomy, which is obtained as explained in Section 3. We will base our approach on the well known *principal-agent* model (Laffont and Martimort 2002; Caillaud and Hermalin 2000) from economics, in which a principal (a service requester) requests an agent (the provider) to perform a specific task.

The outcome of the task execution affects the principal's utility, who will therefore be interested in influencing the efforts that the agent puts in performing the task. Efforts are expressed in terms of available actions, which have associated execution costs. In the so-called *hidden action* setting (Caillaud and Hermalin 2000), it is assumed that the actual actions as executed by the agent are unobservable to the principal. Instead, only some performance measures of such actions are observed. Actions determine, usually stochastically, the obtained performance. Performance is therefore a random variable whose probability distribution depends on the actions taken by the agent. This stochastic nature captures the fact that there are externalities in the environment that the agent has no control over. The principal will therefore want to establish an incentive schedule in order to encourage the agent to choose the actions better leading to an intended performance standard. In our approach, we slightly change the model by putting forward a new entity – an *incentive policy maker* in charge of creating and applying incentive schedules to keep agents conforming to different standards. In other words, it is not the principal (consumer of the service), but this policy maker who sets and applies incentives to the agents (providers).

4.1 Targeting standards

As described in Section 3, standards are generated through the use of an averaging function applied to task execution outcomes of a group of provider agents that have been clustered within a specific role. Since, according to our model, standards allow requesters to identify expected values for the outcomes of tasks when executed by a specific provider, we consider a standard as a target that agents should meet. Any deviation from the standard is considered as a sub-optimal outcome. Figure 1 illustrates this notion, where ζ represents the target standard that the requester

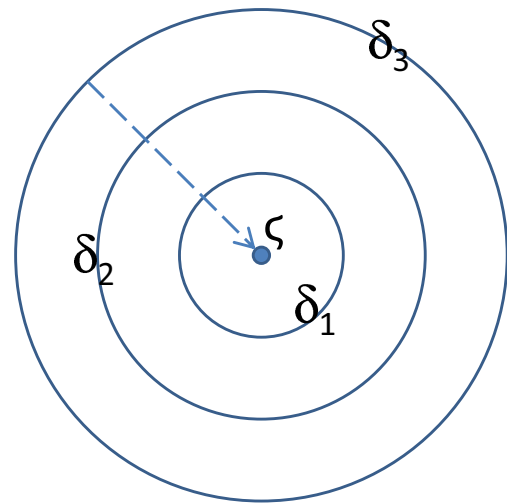


Fig. 1 A standard as a target

would expect, and each concentric circle labelled with a δ_i denotes equidistant performances to the target. These concentric lines highlight the fact that we shall consider deviations in any direction (left or right, upwards or downwards) to be equally harmful in terms of expected values. The arrow pointing towards the centre discloses the aim of our incentive-based approach, with which we will try to encourage providers to better target the standard.

4.2 Actions and outcomes

In our model, we will assume that each provider has a set of actions at its disposal, each with a cost and a probability function for obtaining different performance outcomes. Following a finite model for actions and outcomes, we have that:

Definition 2 The provider has a set of possible actions $\mathcal{A} = \{a_1, \dots, a_n\}$ at its disposal, each having an associated cost, denoted by $Cost(a_i)$.

Definition 3 The possible observable outcomes that the provider may obtain is an ordered set $\mathcal{X}_{attr} = \{x_1, \dots, x_m\}$.

Note that \mathcal{X}_{attr} is the set of possible values for measuring the performance of the task being evaluated at the attribute *attr*. For the sake of simplicity, from now on we take into consideration only one of the attributes of the task, in order to minimise the complexity in notation. Then, \mathcal{X} refers to a \mathcal{X}_{attr} for whatever attribute we are evaluating in the task.

Definition 4 There is a probability distribution function for \mathcal{X} given an action in \mathcal{A} , where $p(x_k|a_i)$ is the probability of

obtaining outcome $x_k \in \mathcal{X}$ when performing action $a_i \in \mathcal{A}$. We have that $\sum_{k=1}^m p(x_k|a_i) = 1$, for all $i \in [1, n]$, where m is the number of possible outcomes and n is the number of actions.

4.3 Incentives

Given the fact that only outcomes, and not efforts, are observable to the principal, incentives are specified through an *incentive schedule* mapping possible outcomes to incentive values to be collected by the provider, according to Definition 5.

Definition 5 An *incentive schedule* $I : \mathcal{X} \rightarrow \mathcal{I}$ maps each possible outcome in \mathcal{X} to a specific incentive value in \mathcal{I} .

We look at incentives as producing some change in the utility the agent would obtain by showing its natural behaviour if no incentives were in place. In this sense, $\mathcal{I} = \{\iota : \iota \in [-1, 1]\}$, where positive values denote percentage increases in utility and negative values denote percentage decreases in utility. When $\iota = 0$ there is no incentive in place. Therefore, positive incentives are considered as rewards for agents to foster the performance of the actions the incentive is applied to, while negative incentives represent an attempt to discourage agents from performing non-desired actions.

4.4 Providers decision rationale

Based on the stochastic model of action outcomes explained above, each provider is assumed to be an expected utility maximizer agent. Therefore, when choosing the action a

to perform it will seek to maximize expected utility (Von Neumann and Morgenstern 1980):

$$\arg \max_{a \in \mathcal{A}} \mathbb{E}_a = \sum_{i=1}^m [p(x_i|a) \cdot u(x_i, I(x_i))] - Cost(a) \quad (1)$$

where $u(x_i, I(x_i))$ is the utility the agent gets from obtaining performance outcome x_i , taking into account the incentive $I(x_i)$ it will get from such a performance. We define function $u(\cdot, \cdot)$ as follows:

$$u(x, \iota) = u(x) \cdot (1 + sens(\iota)) \quad (2)$$

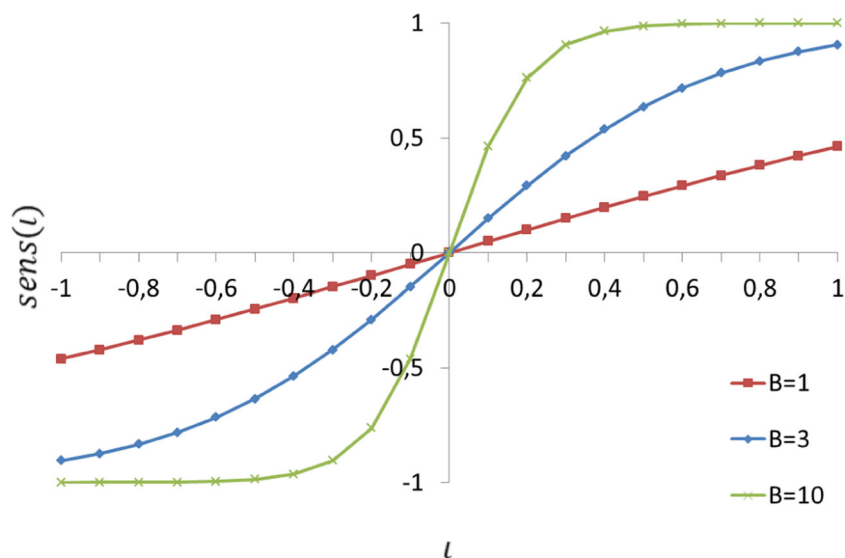
This function encompasses two sub-functions: the prior utility $u(x)$ collected according to the outcome x obtained, and the effect on this utility of the incentive value ι applied. We model such an effect with a sensitivity function $sens : \mathcal{I} \rightarrow [-1, 1]$, which translates an incentive value to its actual perceived impact on the utility of the agent:

$$sens(\iota) = \frac{2}{1 + e^{-\iota B}} - 1 \quad (3)$$

Parameter $B \in \mathbb{N}^+$ allows us to tune the sensitivity of the agent with respect to incentives: higher B values make the agent more sensitive to incentives, while with lower ones the agent will tend to behave the same regardless of any incentives.

Figure 2 shows some examples of Eq. 3 to model $sens(\iota)$ with different B values. When providers are not offered any incentive, they simply obtain the prior utility $u(x)$ as a result of Eq. 2. Negative incentives (punishments) diminish the utility of the provider, whilst positive incentives increase it.

Fig. 2 Different curves for Eq. 3, varying B



5 Maintaining performance standards through incentives

Given the previous performance of each provider, on which standards (via roles) have been defined (as described in Section 3), it may be the case that agents deviate from the standard they were able to meet before. This is due to the evolving nature of the environment in which the agent operates. Since agents are expected utility maximizers, their decision regarding which action to employ when executing a task is conditioned by a number of factors, which we can identify by analysing Eq. 1. Any changes in these factors are thus possible causes for a deviation from the standard characterizing each agent's assigned role:

1. Costs of the actions agents have at their disposal;
2. Effectiveness of available actions, that is, their probability distributions over performance outcomes;
3. Prior utilities that agents get from obtaining each possible outcome;
4. Sensitivity of agents with respect to any incentives they may be offered.

Note that changes in sensitivity are only relevant when there are already incentives in place. Changes in action costs may lead the agent to apply less costly actions, whose outcomes may be different. Changes in the effectiveness of actions may be due to environmental factors not under the control of the agent. In this paper, we assume agents somehow become aware of changes in any of these factors in order to take them into account when deciding which actions to perform. We can easily think of external factors causing these changes. For instance, in a supply chain, fluctuations on prices for different inputs (e.g. parts or raw materials obtained from suppliers) will certainly influence the cost of executing the task. As for outcome probabilities, the agent may be able to update these estimations on-line, according to run-time experience.

These deviations in performance render the role clustering (obtained as described in Section 3) unfit to represent the current performances of agents in the system, in terms of the standards extracted from the roles. Therefore, in order to maintain role stability when agents deviate from agreed standards, the system may determine and employ an appropriate incentive schedule $I : \mathcal{X} \rightarrow \mathcal{I}$ (see definition 5). Since actions are not observable, this schedule is based exclusively on the measurable outcomes of task execution, which for the sake of defining appropriate incentive schedules are compared with the target outcomes characterizing the roles. The *incentive policy maker* (IPM) does not have access to the factors influencing the agents' decision making as this is considered to be private information.

The goal of the IPM is to keep on target the agents playing a specific role, i.e., agents should obtain outcomes

as close as possible to the target outcome of the role. We assume the IPM prefers to achieve this aim with the least incentives needed. In case of failure to accomplish this aim, or if by doing so the IPM has to apply a too costly incentive schedule, then it is time to somehow reorganize the agents that are seen as no longer being able to perform the role at a bearable cost. This is the topic of Section 6.

5.1 Cost-effective incentive schedules

Given an incentive schedule offered to the agents playing a specific role, we may determine its effectiveness by looking at the outcomes that are obtained once that schedule is in place. We should also take into account the cost of applying the incentive schedule. Given the stochastic nature of agent efforts in terms of obtained outcomes, an incentive schedule's effectiveness will typically oscillate around some value, regardless of there being any changes in the environment that lead agents to change their chosen actions. For this reason, in order to compute an incentive schedule's quality $Q(I)$, we aggregate a sequence $X = \langle x^1, x^2, \dots, x^{n_o} \rangle$ of n_o obtained outcomes ($x^i \in \mathcal{X}$), and compare them with the target outcome x^* . We define $Q(I)$ as:

$$Q(I) = \omega \cdot \text{targetHit}(X) - (1 - \omega) \cdot \text{totalCost}(I, X) \quad (4)$$

$$\text{targetHit}(X) = n_o - \sum_{i=1}^{n_o} |x^i - x^*| \quad (5)$$

$$\text{totalCost}(I, X) = \sum_{i=1}^{n_o} |I(x^i)| \quad (6)$$

The *total cost* of the incentive schedule takes into account actually paid incentives, which depend on the outcomes obtained. By using the modulus of the incentive we seek to give the same weight to paid or collected incentive values – without the modulus the IPM would tend to prefer penalizing providers as opposed to paying them incentives or to simply stand still. We thus have that $\text{totalCost}(X) \in [0, n_o]$. On the other hand, *target hit* measures the incentive schedule's effectiveness in inducing agents to meet the target. Any values outside the target are seen as deviations that need to be minimized in terms of role maintenance – for simplicity we assume $\mathcal{X} \subset [0, 1]$, which entails $\text{targetHit}(X) \in [0, n_o]$. Combining these two functions, we have $Q(I)$ to be within the range $[(\omega - 1) \cdot n_o, \omega \cdot n_o]$. Factor $\omega \in [0, 1]$ allows us to balance the relative importance of these two conflicting goals, e.g., by giving priority to obtained performance over how much it costs to achieve it in terms of incentives paid.

5.2 Search space

Incentive schedules specify, for each $x \in \mathcal{X}$, an incentive value $\iota \in \mathcal{I}$. We can therefore represent an incentive schedule as a vector $\iota = [\iota_1, \dots, \iota_m]$, where $m = |\mathcal{X}|$ is the number of possible observable outcomes (see definition 3) and each $\iota_i \in \mathcal{I}$. In the quest to find out the best incentive schedule, measured both in terms of effectiveness and cost, we need to reduce the search space for the IPM, e.g. by limiting the search to incentive schedules composed of values within the set $[\mathcal{I} \cdot 10] / 10$, which gives us discrete incentive values with 0.1 steps.

Depending on the number of outcomes to consider, this may still give us a huge number of schedules to experiment with. We can slightly alleviate this issue by taking into account the intuitive heuristic that we should promote outcomes closer to the target no less than outcomes farther away. Using this principle, the number of incentive schedules available is given by

$$C_{|\mathcal{X}|}^{|\mathcal{I}|+|\mathcal{X}|-1} = \frac{(|\mathcal{I}| + |\mathcal{X}| - 1)!}{(|\mathcal{I}| - 1)! |\mathcal{X}|!}$$

Table 1 shows the number of incentive schedules according to different sizes of the outcomes set, and taking $|\mathcal{I}| = 21$ (which is the size of set $[\mathcal{I} \cdot 10] / 10$ when considering $\mathcal{I} = \{\iota : \iota \in [-1, 1]\}$). As we can see, even when considering a small number of outcomes, the number of incentive schedules is quite large, growing exponentially as $|\mathcal{X}|$ increases.

Notice that applying a single incentive to the target outcome may comprise a suboptimal solution. On one hand, the IPM does not know how precise are the actions chosen by the agents, which means that the action most likely obtaining the target outcome may still be quite noisy. On the other hand, cheaper incentive schedules may be found by taking into account a combination of incentives to different outcomes.

Table 1 Number of incentive schedules for varying $|\mathcal{X}|$

$ \mathcal{X} $	Number of incentive schedules
2	231
3	1771
4	10626
5	53130
6	230230
7	888030
8	3108105
9	10015005
10	30045015

5.3 Finding appropriate incentive schedules

Unlike typical approaches in game theory, we do *not* assume that agents' decision variables (action costs, their probability distributions over outcomes, or utility functions on those outcomes and any employed incentives) are known to the incentive policy maker. We therefore need to go through the search space of possible incentive schedules in order to find the ones that prove to be more cost-effective, by actually trying them out. Given the high number of schedules to experiment with, some heuristics are needed to guide the search.

In the following sections we introduce three approaches for searching for an appropriate incentive schedule. An important feature of these approaches is that they are meant to work on-line: the IPM will be searching for the most cost-effective incentive schedule, while at the same time trying to maximize the accumulated quality of the incentive schedules that are actually employed.

We should mention at this time that it is not our purpose to propose a best alternative in terms of search strategies, but instead to compare a few approaches and see how they fare in terms of some measurable criteria. One such criterion is related with how each approach is able to avoid reallocation of agents to other roles (which is further explained in Section 6).

5.3.1 On-line local search

Given the high number of schedules to experiment with, in this section we follow a *local search* approach to seek an optimal incentive schedule. More specifically, we employ a hill-climbing procedure, by successively trying to find out neighbouring incentive schedules that are better than the currently employed one. In order to find them, however, we need to try out incentive schedules before we know how worthy they are, which makes the search more stochastic. Furthermore, given the dynamics of the environment, these quality values are not constant over time, and thus exploration must always be an option once a change is detected in the environment.

One crucial aspect of local search algorithms is the definition of the neighbourhood function. To generate the neighbours of an incentive schedule, we introduce a step change (upwards or downwards) in the incentive value being applied to any of the outcomes. We then correct the schedule obtained so that outcomes closer to the target have at least the same incentive as outcomes farther away (as mentioned in Section 5.2). This gives us a cardinality of at most $2 \cdot |\mathcal{X}|$ in the set of neighbours of each possible schedule.

The local search procedure is illustrated in Algorithm 1. At each step, we start by checking if we are applying the incentive schedule that is known to be the best (line 2):

if yes, we update it (line 6). If we are applying a different incentive schedule (lines 7-10), we compare the current schedule with the best known (line 7), and update if it is better (lines 8-9). Then we randomly explore, with probability $1 - e^{(||bestQ||-1) \cdot \tau}$ (line 11), the neighbours of the best schedule (line 12): $||bestQ||$ is the normalized value for *bestQ* to the range [0, 1], and τ is a temperature parameter. The better the schedule is, the less likely we will explore, exploiting instead the best schedule we know of (line 14). When a significant change is detected in the environment, measured in terms of a decrease in the quality of the best known schedule (line 3), we promote exploration by resetting the temperature τ (line 4); τ is then decayed in every step (line 16) according to the time we allow the search to proceed (see Section 6), until it reaches nearly 0 (determining no exploration).²

In order to compute the quality of each incentive schedule we need to employ it sufficient time to aggregate new evidence to fill in sequence *X* (see Eq. 4).

Algorithm 1 On-line local search for incentive schedules

Require: *best* {the best known incentive schedule}
bestQ {the previously known quality of *best*}
current {the currently applied incentive schedule}
 τ {a temperature}

```

1: loop
2:   if current = best then
3:     if detectChange(Q(current), bestQ) then
4:       resetTemperature( $\tau$ )
5:     end if
6:     bestQ  $\leftarrow$  Q(current)
7:   else if Q(current)  $\geq$  bestQ then
8:     best  $\leftarrow$  current
9:     bestQ  $\leftarrow$  Q(current)
10:  end if
11:  if random <  $1 - e^{(||bestQ||-1) \cdot \tau}$  then
12:    current  $\leftarrow$  RandomNeighbour(best)
13:  else
14:    current  $\leftarrow$  best
15:  end if
16:  decayTemperature( $\tau$ )
17: end loop

```

5.3.2 On-line tabu search

The search space we are dealing with is quite plateaux-like, given the fact that many neighbours of a given schedule will have the same exact quality. This is a challenge for an approach based on hill-climbing, such as the one presented

in Section 5.3.1. One well-known technique to tackle such kind of search spaces is tabu-search, which we explore in this section. Tabu-search includes a list of forbidden nodes (schedules) with the aim of escaping local-optima.

Algorithm 2 shows our approach. The *tabulist* contains a small number of the most recently visited incentive schedules; whenever we add an element to this list (line 13), if the number of elements exceeds its size we discard the most outdated one (in a first-in-first-out fashion). Besides keeping the best schedule found so far, we also keep the point from which we will continue the search, by following one of its neighbours that are not in the tabu-list (line 12). The remaining parts of the algorithm are similar to Algorithm 1.

Algorithm 2 On-line tabu search for incentive schedules

Require: *best* {the best known incentive schedule}
bestQ {the previously known quality of *best*}
current {the currently applied incentive schedule}
 τ {a temperature}
tabulist[] {a size-limited list of recently visited incentive schedules}
searchpoint {the incentive schedule to search from, initially *current*}

```

1: loop
2:   if current = best then
3:     if detectChange(Q(current), bestQ) then
4:       resetTemperature( $\tau$ )
5:     end if
6:     bestQ  $\leftarrow$  Q(current)
7:   else if Q(current)  $\geq$  bestQ then
8:     best  $\leftarrow$  current
9:     bestQ  $\leftarrow$  Q(current)
10:  end if
11:  if random <  $1 - e^{(||bestQ||-1) \cdot \tau}$  then
12:    neighbour  $\leftarrow$  NontabuRandomNeighbour(searchpoint, tabulist)
13:    tabulist.add(searchpoint)
14:    current  $\leftarrow$  neighbour
15:    searchpoint  $\leftarrow$  neighbour
16:  else
17:    current  $\leftarrow$  best
18:  end if
19:  decayTemperature( $\tau$ )
20: end loop

```

5.3.3 Reinforcement learning

Learning on-line, i.e. by interacting with the environment and obtaining appropriate rewards, is the aim of *reinforcement learning* (RL) (Sutton and Barto 1998). In this section, we look at the problem of searching for an optimal incentive schedule as a reinforcement learning problem. More specifically, the setting we are addressing is similar to an *n-armed bandit problem* (Sutton and Barto 1998). The incentive policy maker (the *learner*) will try to determine, by exploring its action³ set, the best possible *incentive schedule I* whose

²Note that this temperature mechanism is not related to simulated annealing, in which the temperature determines the probability of choosing a worse solution; in Algorithm 1 we use it simply to induce and then to reduce exploration, while a new schedule will only be kept if it is found to be better than the best we know of (hence the hill-climbing flavour).

³We emphasize that these are the learner's actions (i.e., those available to the IPM), and not the actions of the provider agent as discussed in Section 4.2. We here use the same term *action* because it is well established in RL literature.

quality $Q(I)$ is updated in terms of a *reward* obtained from the environment.

The problem faced by the learner is a continuing task: given the sources for deviating behaviours identified in Section 5, providers may decide differently when facing a specific incentive schedule at different times, which makes the problem non-stationary. Reinforcement learning naturally encompasses this kind of situations by maintaining a trade-off between exploitation (taking advantage of the actions that have been found as good) and exploration (trying out other actions whose effect is not totally known or up-to-date). However, the main problem with a RL approach in our case is related with the size of the action space, as shown in Table 1.

Algorithm 3 Reinforcement learning for incentive schedules

Require: $qv(A)$ {estimated Q values for incentive schedules, initially empty}
 $best$ {the best known incentive schedule}
 $bestQ$ {quality of $best$ }
 $current$ {the currently applied incentive schedule}
 α {a step-size parameter}
 τ {a temperature}

```

1: loop
2:    $qv(current) \leftarrow qv(current) + \alpha \cdot (Q(current) - qv(current))$ 
3:   if  $current = best$  then
4:     if  $detectChange(qv(current), bestQ)$  then
5:        $resetTemperature(\tau)$ 
6:     end if
7:      $bestQ \leftarrow qv(current)$ 
8:   else if  $qv(current) \geq bestQ$  then
9:      $best \leftarrow current$ 
10:     $bestQ \leftarrow qv(current)$ 
11:   end if
12:    $AddMissingActions(qv, Neighbourhood(current), qv(current))$ 
13:   if  $random < 1 - e^{(||bestQ||-1) \cdot \tau}$  then { $\epsilon$ -greedy action selection}
14:      $current \leftarrow RandomAction(qv)$  {explore}
15:   else
16:      $current \leftarrow best$  {exploit}
17:   end if
18:    $decayTemperature(\tau)$ 
19: end loop

```

Our approach is shown in Algorithm 3. Given the observed reward $Q(\cdot)$ (obtained as in Eq. 4), we update the estimated Q value $qv(\cdot)$ for the incentive schedule being applied (line 2). Note that, unlike the previous algorithms, $Q(\cdot)$ is here taken as a reward that is used to update the current estimate on the true value of the schedule being used. We then update the best schedule in a way similar to the previous algorithms (lines 3-11). As in the previous approaches, when a significant change is detected in the environment we reset the temperature τ (lines 4-6), which will be later subject to decay (line 18).

Given the size of the action space, we take an incremental approach to the number of actions to consider: schedules

that are neighbours of the current one are added to the action space (if not already there), where for bootstrapping they are initialized with the Q value estimate of the current schedule (line 12). Finally, an ϵ -greedy action selection policy determines that we explore with probability $1 - e^{(||bestQ||-1) \cdot \tau}$ a random action (lines 13-14), or else we stick with the best one (line 16).

6 Re-organization

Given the organizational flavour of role taxonomies, we have preferred an incentive-based policy to a reorganization of the system in terms of up-to-date skills of agents. This preference is due to both practical and fundamental reasons. The computational complexity of creating or updating a role-taxonomy may be considerable, especially when little is known about the agents that are to take part in this organization, and thus about the roles that are to be created (see e.g. the clustering approach in Hermoso et al. (2010)). Furthermore, the very notion of roles as a descriptive facet of an organization presumes some notion of stability, as roles provide a means to identify specific agents within the organization. If we reorganize too often, this sense of long-term existence is lost.

Nevertheless, there will be situations in which reorganization is a better option as compared to applying incentive schedules with the aim of influencing the providers' behaviour. On one hand, although an effective incentive schedule may be found, applying it may bear a significant cost. On the other hand, there will be situations in which an effective incentive schedule is not found. This may happen either because the search procedure employed is not able to find it (local search approaches are typically vulnerable to local optima) in a reasonable time (e.g. due to a very large search space), or because environmental changes have reduced the influencing ability of the IPM, therefore lowering the effectiveness of the most effective incentive schedule available.

Having this in mind, we add a "reorganization" ability to the system. In our setting, reorganizing aims at increasing the cohesion of the roles in terms of the skills of the agents to which those roles have been assigned. As mentioned above, this reorganization step is taken to be costly. However, instead of using a more thorough approach such as in Hermoso et al. (2010), in this paper we assume that we are able to assess the role that each agent is able to perform best, without any incentives. In that sense, to reorganize means to reassign an agent to that role.

Algorithm 4 expresses the policy that we follow to decide when to reallocate an agent. In order to assess if we should reallocate an agent, we must take into consideration its most recent outcomes ($X[]$), as well as the quality of the

most recently applied incentive schedules ($Q[]$). These values will be updated while the agent executes tasks, and while the IPM tries to find out appropriate incentive schedules, respectively. Given the continuous search for better incentive schedules,⁴ the stochastic nature of outcomes, and possible environmental changes, both outcome and Q values may oscillate. For this reason we must consider a set of recent values for these two variables.

In Algorithm 4, we start by checking if the outcomes obtained are significantly off-target, or if the average Q values are below a given threshold δ (line 2). If so, we either remember the point in time where we have detected this agent's "default" (line 4), or if detected before we determine if sufficient time (at least Δ) has elapsed for the IPM to find an appropriate incentive schedule (line 5) – this is the case when we choose to reallocate the agent (lines 6-9). Reallocating the agent entails that the agent will have a new target outcome. In case the agent is not deviating, we simply forget its previous default time, if any (line 12).

Algorithm 4 Reallocation policy

Require: *targetOutcome* {the target outcome for the agent}
 $X[]$ {recently obtained outcomes}
 $Q[]$ {recently obtained Q values, as per Eq. 4}
defaultTime {the time when the agent has defaulted}
 δ {the default threshold}
 Δ {the time allowed for searching for incentive schedules}

```

1: loop
2:   if OffTarget( $X[], targetOutcome$ ) or Avg( $Q[]$ ) <  $\delta$  then
3:     if defaultTime = null then
4:       defaultTime  $\leftarrow$  now
5:     else if now – defaultTime >  $\Delta$  then
6:       targetOutcome  $\leftarrow$  Reallocate()
7:        $X[] \leftarrow \emptyset$ 
8:        $Q[] \leftarrow \emptyset$ 
9:       defaultTime  $\leftarrow$  null
10:    end if
11:  else
12:    defaultTime  $\leftarrow$  null
13:  end if
14: end loop

```

The two parameters δ and Δ allow us to adjust the implicit cost of reorganization, as compared to that of search. By increasing the value of the default threshold δ , we demand for more cost-effective incentive schedules to prevent reorganization. Lower δ values, on the other hand, will allow for less cost-effective schedules, meaning that we consider reorganization as being a more costly operation to undertake. Similarly, a higher time frame Δ will allow the search algorithm employed by the IPM to have more

time, while a lower Δ value will tolerate less deviations to occur.

7 Evaluation

We have used Repast Symphony as a simulation framework to evaluate our approach. In this section, we present the results of different experiments carried out to demonstrate the suitability of the different methods for maintaining the providers attached to the standards. We have designed experiments to present how our proposal learns appropriate incentives for different sets of providers, by using the learning and optimization approaches put forward in Section 5.

7.1 Provider setup

First we need to present how we set up the providers for the experiments. We will work with populations of 100 providers. We initialize each provider as follows:

- *Effort-Cost*. The costs of efforts (see definition 2) are initialized randomly following a uniform distribution within $[0, 1]$.
- *Effort-Outcome*. We need to endow providers with a method for generating possible outcomes in \mathcal{X} from different chosen efforts (see definition 4). We model this using beta distributions, in order to allow for stochastic outcomes, avoiding a deterministic provider behaviour. The shape of this type of functions is controlled by two parameters, α and β . We set $\alpha = 1 + (c * p - c)$ and $\beta = p - (c * p - c)$, where c is a centre value (corresponding to one of the outcomes in \mathcal{X}) and p is a peak factor. We set c randomly with a uniform distribution within \mathcal{X} , and p randomly with a uniform distribution in the range $[100, 200]$. As mentioned in Section 5.1, we assume $\mathcal{X} \subset [0, 1]$; we set $|\mathcal{X}| = 7$.
- *Outcome-Utility*. The prior utilities (see Eq. 2) for different outcomes that may be obtained are initialized randomly following a uniform distribution within $[0, 1]$.

In all experiments presented in this section, we have set $B = 10$ for Eq. 3.

7.2 Experimental dynamics

Once we have clarified how the population of providers is set up now we proceed to explain the functioning of experimental scenarios. We will run every experiment 1000 time steps. In each time step, every provider will be required to perform an action. The provider will, naturally, choose the action that maximises its expected utility (following Eq. 1).

⁴Except for very low temperature values, see Algorithms 1-3.

In order to test how different methods for learning incentive schedules perform, we introduce changes at ticks 100 and 600. Changes are applied over 50% of the providers. We show results for two types of scenarios:

- *Soft changes.* This scenario tests the ability of the search methods to find appropriate incentive schedules when smooth changes happen. In both changing ticks, changes consist of applying variations on the costs of each effort the provider has at its disposal: $Cost(a_i) \pm 0.25$. Increasing or decreasing the cost is modelled as flip-a-coin stochastic process. The other two functions that define the provider behaviour remain unchanged.
- *Random changes.* This scenario tests the robustness of the learning methods against severe changes in the providers' population. In both changing ticks, we modify providers' effort costs, outcome utilities and effort-outcome beta distributions in a random way. That is, we randomly re-initialise costs and prior utilities, and set new random centres and peak values for the beta distributions.

7.3 Configuration of search and reorganization algorithms

We will denote each search algorithm as follows: LS (on-line Local Search), TS (on-line Tabu Search) and RL (Reinforcement Learning); NS represents a No Search approach, i.e., no incentives will be used.

Regarding the different parameters for the search and reallocation algorithms we have set these up as follows:

- For the reallocation policy (see Algorithm 4), $\Delta = 300$ (the mechanism allows a time window of 300 time steps to let the search method find out suitable incentives), and $\delta = 0.8$ (average Q values below 0.8 will be seen as defaults).

- For RL, we set $\alpha = 0.8$ (see Algorithm 3), allowing the algorithm to quickly adapt to changes.
- For LS, TS and RL, the value to reset the temperature τ is 10.
- For calculating $Q(\cdot)$ values, we use $\omega = 0.7$ (see Eq. 4), therefore giving priority to achieve the target instead of to the associated costs.

These values were chosen after some empirical testing.

7.4 Metrics

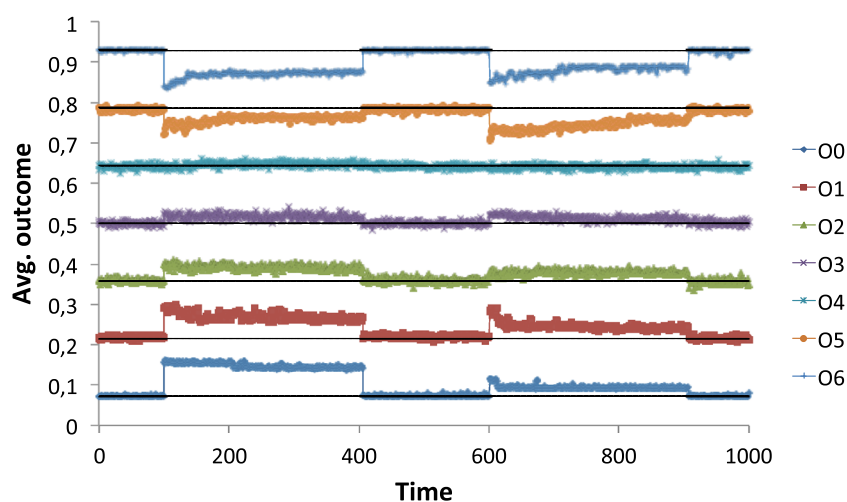
We use two different metrics to interpret the obtained results. We use the average $Q(\cdot)$ obtained by the agents in every time step. This is calculated as described in Eq. 4.

On the other hand, we are also interested in measuring how good different methods are in terms of the number of reallocated providers. The lower this number the better the method performs, since this represents a better learning of incentives, so saving a costly re-organisation process, as we pointed out in Section 6.

7.5 Results

In order to explain the functioning of the different methods in terms of obtained outcomes we present in Fig. 3 the average outcomes obtained by providers over time when LS tries to learn incentives. In that figure we observe the different 7 outcomes we define this particular experiment, as well as the standard values (black straight lines) for each of them. Marked and in colour are the different average outcomes of the providers. That is, every O_i line denotes the average outcome of the providers allocated in a particular target outcome. From the curves, we can notice how average outcomes deviate from the target outcomes, since a

Fig. 3 Functioning of LS in terms of average outcomes for random changes scenario



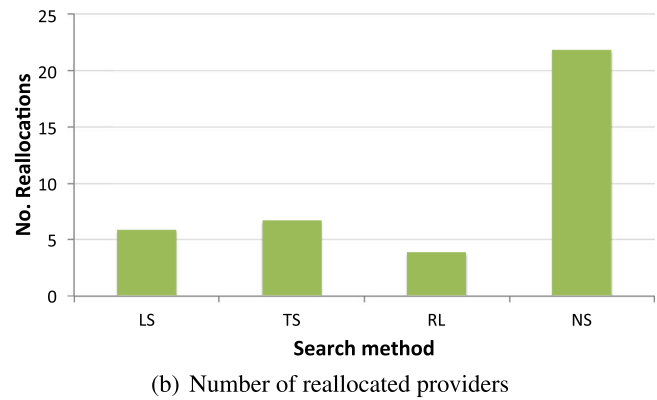
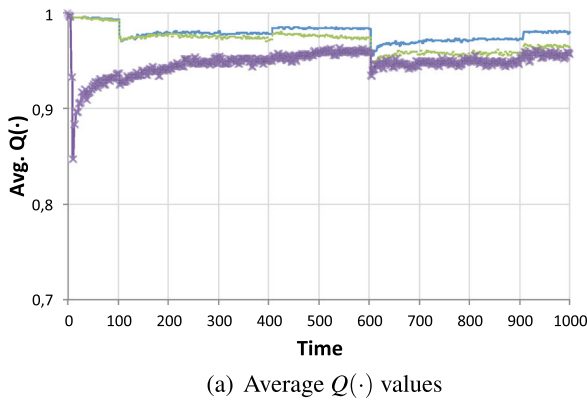


Fig. 4 Evaluating results of different search methods for soft changes

significant number of providers change their behaviour. As we have mentioned, changes occur in the system at time steps 100 and 600, so between these times, the learning method should be able to keep part of the providers back to the target outcome. This process is observed after every change since the average outcomes tend to drift back to the standards over time. In those cases in which incentives are not effective, re-organisation is needed and providers are reallocated to other target outcomes (around ticks 400 and 900).

Figure 4 shows the results for the scenario in which soft changes are applied. Figure 4b depicts the result of each method in terms of the number of reallocated providers, together with the absence of incentives for comparison purposes. As expected, the three approaches we propose (LS, TS and RL) greatly outperform the case with no search (less reallocations needed), which illustrates the possibility of using incentive schedules to keep agents on track at a bearable cost.

Following Fig. 4a, we can see that the proposed methods allow to quickly find out proper incentives to keep most providers behaving according to standards. The population-change ticks are clearly observable, while the reallocation

ones are less so, indicating the apparent success in LS, TS and RL. The method obtaining the best $Q(\cdot)$ values is LS, followed closely by TS. The online nature of RL, trying to explore the search space from the beginning, makes it less cost-effective in this scenario, where initially incentive schedules are not needed at all.

Figure 5 shows the results for the scenario in which random changes are applied. In Fig. 5b, we observe the result of the methods in terms of the number of reallocated providers. The three approaches still outperform the case with no search. However, there still exists a high number of providers that must be reallocated in this case, because the search algorithms are not able to find incentive schedules that are cost-effective enough to prevent reallocations. In fact, in this scenario the inflicted changes may become so drastic that could make the methods useless, since even extreme incentives may be short in preventing providers from misbehaving (it may be even the case that the random change has made it impossible for the provider to obtain the target outcome no matter what).

Figure 5a confirms the hard task of finding cost-effective incentive schedules: crisp changes are noticed after the search period, when agents are reallocated to appropriate

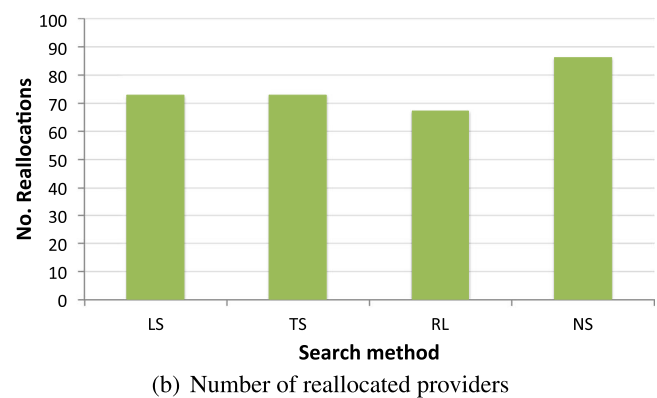
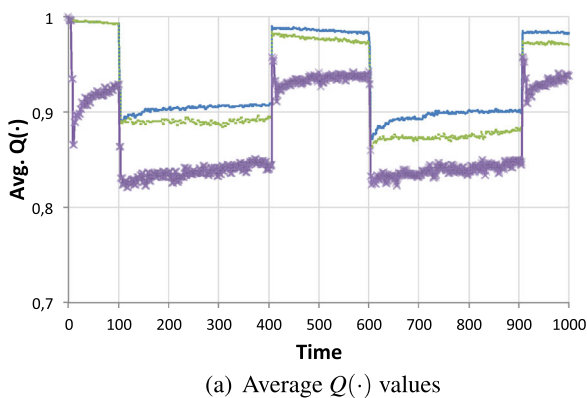


Fig. 5 Evaluating results of different search methods for random changes

roles. As in the previous case, LS is the method obtaining the best $Q(\cdot)$ values.

8 Discussion

In this paper, we have put forward a theoretical approach to build standards from roles in dynamic task-oriented MAS. Although the creation of standards has been more deeply studied in the fields of economics and finance, in the MAS community there have been some attempts to dynamically build social structures to foster interactions. For instance, there are many approaches on how norms are formed and how they emerge from expectations. In Winikoff and Cranefield (2009), the authors present a work that gathers users expectations for social interactions to transform them into logic formulae that can be used in order to check an eventual outcome. The main difference with our approach is that they use explicit requests to the users to gather their expectations, while we automate that process by using the role creation mechanism. Other approaches related to this issue are Castelfranchi et al. (2003) and (Conte and Castelfranchi 1999), in which the authors discuss how prescriptions might emerge from individual expectations, eventually forming norms.

There are economic approaches also founded on the emergence of standards, such as Sherstyuk (2000), in which Sherstyuk proposes a method to set an appropriate performance standard to develop optimal contracts, i.e., contracts in which the provider agent's best choice is to keep the standard through its action. In this paper, however, we are not interested in obtaining optimal performance standards; instead, we are concerned about how to maintain the level of those standards once they have been created.

Along the same lines, (Centeno et al. 2011) present an approach on adaptive sanction learning by exploring and identifying individuals' inherent preferences without explicit disclosure of information. That is, the mechanism learns over which attributes of the system should modifications be applied in order to induce agents to avoid undesired actions. In our case, we adhere to a more formal scenario, in which interactions are regulated by means of some kind of contracts and, besides, we assume that attributes that may be modified by means of incentives are already known by the mechanism.

The approach taken in Lopes Cardoso and Oliveira (2011) also assumes that the mechanism knows which attributes it should tweak in order to influence agents' behaviours, namely by adjusting deterrence sanctions applicable to contractual obligations agents have committed to. The notion of social control employed there is similar to our notion of role standard maintenance, although instead of a run-time discovered standard, a fixed threshold is used

to guide the decisions of the policy maker. Moreover, only sanctions (seen as fines) are employed to have a deterrence effect of discouraging agents from misbehaving, while here we are more interested on incentivising agents to do their best while executing the tasks they are assigned to.

In fact, in the model and experiments we have reported in this paper we encompass both incentives (positive rewards) and penalties (negative sanctions) – both dubbed incentives in definition 5. It may be arguable whether agents in an open system will be willing to submit themselves to sanctions for the benefit of a service provided by the open system's infrastructure. That is, which other assumptions must be in place so that this scenario is reasonable? Our guess is that this will depend on domain-dependent issues which are out of the scope of this paper. In any case, the theoretical approach that we have proposed remains untouched if we find it more reasonable to consider proper incentives (i.e. positive rewards) only. This will however have an impact on the performance of the search approaches we have designed and experimented with.

One could argue that the proposed incentive mechanism is exploitable by providers, once they notice that by intentionally under-performing they are able to get access to incentives that would otherwise not be needed. However, such agents are likely to lose in the long run, as soon as the system finds it better to reallocate the agent to another role. Implicitly, we assume that different roles provide different statuses to agents, which has an effect on their private goals or utility.

9 Conclusions and future work

Standards are used as a means to articulate contracts in social interactions. When dealing with organisational multi-agent systems, roles can be used as a reference in order to create standards: a standard can be seen as a measure of the quality of performance of agents playing a certain role. In this paper, we have proposed a mechanism that, on the one hand, creates performance standards from roles discovered at run-time in a multi-agent system and, on the other, provides incentives to make agents maintain a level of performance as close as possible to the standards. When this is not possible, reorganization is the way to eliminate the gap between the descriptive and functional nature of roles and the capabilities of agents.

Although the model is kept mainly theoretical in this paper, which has made the approach more general and domain-independent, we envisage some possible application domains. In manufacturing systems, agents fulfilling different roles when building a craft are supposed to meet and maintain a standard during their work. In social systems such as ruled electronic markets, while standards may

not be known a priori, they can be discovered at runtime and artificially maintained for the sake of the overall market community.

The complex simulation we have implemented provides us a large range of experimentation options that we intend to explore. We believe that it is important to experiment with the incentive schedule learning approach in a changing environment, taking into account the causes for deviations identified in Section 5.

Finally, a more comprehensive framework will embrace the whole cycle from role taxonomy generation, standards measurement, maintenance through incentives, and reorganization. All these steps are needed to address open multi-agent systems, given their dynamic nature.

Acknowledgements The present work has been partially funded by the Spanish Ministry of Education and Science under projects OVAMAH-TIN2009-13839-C03-02 (co-funded by Plan E) and Agreement Technologies (CONSOLIDER CSD2007-0022, INGENIO 2010) and by the Spanish Ministry of Economy and Competitiveness by the project iHAS (grant TIN2012-36586-C03-02).

References

- Bandura, A., & Wood, R. (1989). Effect of perceived controllability and performance standards on self-regulation of complex decision making. *Journal of personality and social psychology*, 56(5), 805–814.
- Boella, G., van der Torre, L., Verhagen, H. (2006). Introduction to normative multi-agent systems. *Computational & Mathematical Organization Theory*, 12(2-3), 71–79.
- Boman, M. (1999). Norms in artificial decision making. *Artificial Intelligence and Law*, 7(1), 17–35.
- Busch, L. (2000). The moral economy of grades and standards. *Journal of Rural Studies*, 16(3), 273–283.
- Caillaud, B., & Hermalin, B. (2000). Hidden action and incentives. Teaching Notes, U.C. Berkeley, accessed at <http://faculty.haas.berkeley.edu/hermalin/agencyread.pdf>.
- Castelfranchi, C., Giardini, F., Lorini, E., Tummolini, L. (2003). The prescriptive destiny of predictive attitudes: From expectations to norms via conventions. In: R., Alterman & D., Kirsh (Eds.) In *Proceedings of the 25th annual meeting of the cognitive science society*, Boston, (pp. 222–227).
- Centeno, R., Billhardt, H., Hermoso, R. (2011). An adaptive sanctioning mechanism for open multi-agent systems regulated by norms. In *Proceedings of the 2011 23rd IEEE international conference on tools with artificial intelligence*, ICTAI '11. IEEE Computer Society, (pp. 523–530).
- Cizek, G.J., & Bunch, M.B. (2007). Standard setting: A guide to establishing and evaluating performance standards on tests. SAGE Publications.
- Conte, R., & Castelfranchi, C. (1999). From conventions to prescriptions. Towards an integrated view of norms. *Artificial Intelligence and Law*, 7(4), 323–340.
- Dignum, V., & Dignum, F. (2001). Modelling agent societies: coordination frameworks and institutions. In: P., Brazdil & A., Jorge (Eds.), *Progress in artificial intelligence: Knowledge extraction, multi-agent systems, logic programming, and constraint solving*, Lecture Notes in Artificial Intelligence, (vol. 2258. Springer, Berlin Heidelberg, pp. 191204).
- Esteva, M., Rodríguez-Aguilar, J.A., Sierra, C., Garcia, P., Arcos, J.L. (2001). On the formal specifications of electronic institutions. In: F., Dignum & C., Sierra (Eds.), *Agent-mediated electronic commerce: The european agentlink perspective*, Lecture Notes in Artificial Intelligence, (vol. 1991. Springer, pp. 126147).
- Fasli, M. (2003). Reasoning about the dynamics of social behaviour. In *Proceedings of the autonomous agents and multi-agent systems conference (aamas '03)*. ACM Press, (pp. 988–989).
- Fasli, M. (2006). On the relationship between roles and power: preliminary report. In *Proceedings of the 2nd computer-aided law and advanced technologies forum (clat 2006) part of the 21st acm symposium on applied computing (sac 2006)*. ACM Press, (pp. 313–318).
- Fornara, N., & Colombetti, M. (2003). Defining Interaction Protocols using a Commitment-based Agent Communication Language. In: *Second international joint conference on autonomous agents and multiagent systems*, ACM, New York, Melbourne, Australia, (pp. 520–527).
- Fornara, N., Vigan, F., Verdicchio, M., Colombetti, M. (2008). Artificial Institutions: A Model of Institutional Reality for Open Multiagent Systems. *Artificial Intelligence and Law*, 16(1), 89–105.
- García-Camino, A., Rodríguez-Aguilar, J.A., Sierra, C., Vasconcelos, W. (2007). Norm-Oriented Programming of Electronic Institutions: A Rule-Based Approach. In: P., Noriega J., Vázquez-Salceda G., Boella O., Boissier V., Dignum N., Fornara E., Matson (Eds.), *Coordination, organizations, institutions, and norms in agent systems II*, LNAI 4386. Springer, Berlin Heidelberg New York, (pp. 177–193).
- Hambleton, R.K., Jaeger, R.M., Plake, B.S., Mills, C. (2000). Setting performance standards on complex educational assessments. *Applied Psychological Measurement*, 24(4), 355–366.
- Hermoso, R., Ramón, B., Billhardt, H., Olger, S., Ossowski, S., Sascha. (2010). Role evolution in open multi-agent systems as an information source for trust. In *9th international conference on autonomous agents and multi-agent systems*. IFAAMAS, (pp. 217–224).
- Hermoso, R., Billhardt, H., Ossowski, S. (2013). Trust-based role coordination in task-oriented multiagent systems. *Knowledge-Based Systems*, 52, 78–90.
- Hubner, J.F., Sichman, J.S., Boissier, O. (2002). A Model for the Structural, Functional, and Deontic Specification of Organizations in Multiagent Systems. In: G., Bittencourt & G. L., Ramalho (Eds.) In *Advances in artificial intelligence – 16th brazilian symposium on artificial intelligence*. Springer, Brazil, (pp. 439–448).
- Jones, A., & Sergot, M. (1996). A Formal Characterisation of Institutionalised Power. *Logic Journal of the IGPL*, 4(3), 427–443.
- Laffont, J.J., & Martimort, D. (2002). In *The theory of incentives: The principal-agent model*. Princeton paperbacks. Princeton: Princeton University Press.
- Lopes Cardoso, H., & Oliveira, E. (2008). Electronic Institutions for B2B: Dynamic Normative Environments. *Artificial Intelligence and Law*, 16(1), 107–128.
- Lopes Cardoso, H., & Oliveira, E. (2011). Social control in a normative framework: An adaptive deterrence approach. *Web Intelligence and Agent Systems*, 9, 363–375.
- López y López, F., & Luck, M. (2003). Modelling Norms for Autonomous Agents. In: E., Chavez J., Favela M., Mejia A., Oliart (Eds.) In *Fourth mexican international conference on computer science*. IEEE Computer Society, Los Alamitos, (pp. 238–245).
- Murphy, K.J. (2000). Performance standards in incentive contracts. *Journal of Accounting and Economics*, 30(3), 245–278.
- North, D.C. (1990). In *Institutions, Institutional Change and Economic Performance*. Cambridge: Cambridge University Press.
- Rehak, M., Gregor, M., Pechoucek, M. (2006). Multidimensional context representations for situational trust. In *Proceedings of*

- the *ieee workshop on distributed intelligent systems: Collective intelligence and its applications*, DIS '06. IEEE Computer Society, Washington, DC, (pp. 315–320).
- Salgé, F. (2005). National and international data standards. In: P.A., Longley M.F., Goodchild D.J., Maguire D.W., Rhind (Eds.) In *Geographical information systems. principles, techniques, management and applications*. 2nd edn. John Wiley, New York, (pp. 693–706).
- Searle, J.R. (1995). In *The Construction of Social Reality*. New York: Free Press.
- Sherstyuk, K. (2000). Performance standards and incentive pay in agency contracts. *Scandinavian Journal of Economics*, 102(4), 725–736.
- Singh, M.P. (1999). An ontology for commitments in multiagent systems: Toward a unification of normative concepts. *Artificial Intelligence and Law*, 7(1), 97–113.
- Southgate, L., Hays, R.B., Norcini, J., Mulholland, H., Ayers, B., Woolliscroft, J., Cusimano, M., McAvoy, P., Ainsworth, M., Haist, S., Campbell, M. (2001). Setting performance standards for medical practice: a theoretical framework. *Medical Education*, 35(5), 474–81.
- Sutton, R., Richard, S., & Barto, A. (1998). In *Reinforcement Learning: An Introduction*. Cambridge: The MIT Press.
- Tavakoli, M., Knapskog, S.J., Herrmann, P. (2008). Cross-situation trust reasoning. In *Proceedings of the 2008 IEEE/WIC/ACM international conference on web intelligence and intelligent agent technology - volume 03*, WI-IAT '08. IEEE Computer Society, Washington, DC, USA, (pp. 67–71).
- Tuomela, R. (1995). In *The Importance of Us: A Philosophical Study of Basic Social Norms*. Stanford: Stanford University Press.
- Urbano, J., Rocha, A.P., Oliveira, E. (2011). In *Transactions on computational collective intelligence v*, (pp. 84–105). Berlin, Heidelberg: Springer-Verlag.
- Von Neumann, J., & Morgenstern, O. (1980). In *Theory of Games and Economic Behavior*, 3rd edn. Princeton: Princeton University Press.
- Walker, A., & Wooldridge, M. (1995). Understanding the Emergence of Conventions in Multi-Agent Systems. In: V., Lesser & L., Gasser (Eds.) In *Proceedings of the first international conference on multi-agent systems*. MIT Press, San Francisco, (pp. 384–389).
- Winikoff, M., & Cranefield, S. (2009). Eliciting expectations for monitoring social interactions. In *Proceedings of the first international conference on computer-mediated social networking*, ICCMSN'08. Springer, Berlin Heidelberg New York, (pp. 171–185).
- Ramón Hermoso** is an assistant professor at the University of Zaragoza. Formerly, he worked as a senior research officer at the University of Essex and as an assistant professor at the University Rey Juan Carlos in Madrid (Spain). He received his PhD from the University Rey Juan Carlos in 2011. His research interests span from trust and reputation mechanism in multiagent systems to the innovation management in social networks. He is author of several publications in journals, books and international conferences, and has participated in more than 15 research projects, funded by both national and international institutions. He has been involved in organising international events and peer reviewing for international journals, conferences and workshops.
- Henrique Lopes Cardoso** obtained his PhD on Informatics Engineering from the University of Porto in 2011. He is an Assistant Professor at the Faculty of Engineering of the University of Porto (FEUP) and a researcher at the Artificial Intelligence and Computer Science Lab (LIACC). He is also a member of the directive board of the Portuguese Association for Artificial Intelligence (APPIA). His research interests include distributed AI, social coordination and regulation of multi-agent systems, adaptive learning agents, and multi-agent systems tools. He has been an active member of relevant European research networks, namely the COST Action IC0801 on Agreement Technologies and the European Network for Social Intelligence (SINTELNET).
- Maria Fasli** is a Professor in the School of Computer Science and Electronic Engineering, University of Essex where she has been a member of staff since 1999. Her research interests lie in agents and multi-agent systems and their theoretical foundations and practical applications, machine learning, analysing and modelling complex data (structured/unstructured), Big Data, as well as semantic-based techniques for user profiling and adaptation including modelling context. She has published in journals and international conferences and specialists workshops in the field of artificial intelligence and multi-agent systems and has participated in international competitions such as the Trading Agent Competition. She has been involved in organising/chairing international events and peer reviewing for conferences and funding organisations. She is the author of *Agent Technology for E-commerce* (Wiley, 2007). In 2006, she was awarded a National Teaching Fellowship by the Higher Education Academy (UK) for her innovations and contributions to learning and teaching.