J Braz Comput Soc (2011) 17:143–161
DOI 10.1007/s13173-011-0035-5

ORIGINAL PAPER

# Recognizing and learning models of social exchange strategies for the regulation of social interactions in open agent societies

**Graçaliz Pereira Dimuro · Antônio Carlos da Rocha Costa · Luciano Vargas Gonçalves · Diego Rodrigues Pereira**

Received: 15 February 2011 / Accepted: 8 July 2011 / Published online: 6 August 2011


**Abstract** Regulation of social exchanges refers to controlling social exchanges between agents so that the balance of exchange values involved in the exchanges are continuously kept—as far as possible—near to equilibrium. Previous work modeled the social exchange regulation problem as a POMDP (Partially Observable Markov Decision Process), and defined the **policyToBDIplans** algorithm to extract BDI (Beliefs, Desires, Intentions) plans from POMDP models, so that the derived BDI plans can be applied to keep in equilibrium social exchanges performed by BDI agents. The aim of the present paper is to extend that BDI-POMDP agent model for self-regulation of social exchanges with a module, based on HMM (Hidden Markov Model), for recognizing and learning partner agents' social exchange strategies, thus extending its applicability to open societies, where new partner agents can freely appear at any time. For the recognition problem, *patterns of refusals* of exchange proposals are analyzed, as such refusals are produced by the partner agents. For the learning problem, HMMs are used to capture probabilistic state transition and observation functions that model the social exchange strategy of the partner agent, in order to translate them into POMDP's action-based state transition and observation functions. The paper formally addresses the problem of translating HMMs into POMDP models and vice versa, introducing the translation algorithms and some examples. A discussion on the results of simulations of strategy-based social exchanges is presented, together with an analysis about related work on social exchanges in multiagent systems.

**Keywords** Social exchange strategy · Recognition and learning of social exchange strategies · Self-regulation of social exchange strategies · Partially observable Markov decision process · Hidden Markov model

G.P. Dimuro · A.C. da Rocha Costa (✉)
Centro de Ciências Computacionais, Universidade Federal do Rio Grande, Av. Itália km 08, Campus Carreiros, 96201-900 Rio Grande, Brazil
e-mail: ac.rocha.costa@gmail.com

G.P. Dimuro
e-mail: gracaliz@gmail.com

L.V. Gonçalves
Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Sul, Núcleo Avançado de Farroupilha, Avenida São Vicente 785, 95180-000 Farroupilha, Brazil
e-mail: llvarga@gmail.com

D.R. Pereira
Programa de Pós-Graduação em Ciência da Computação, Universidade Federal do Rio Grande do Sul, Av. Bento Gonçalves 9500, 91501-970 Porto Alegre, Brazil
e-mail: psador@gmail.com

## 1 Introduction

In Piaget's Theory of Social Exchanges [41], social interactions are seen as service exchanges between pairs of agents, together with the subjective evaluation of those exchanges by the agents themselves, by means of the so-called *social exchange values*: the investment value for performing a service or the satisfaction value for receiving it. The exchanges also generate values of debts and credits that help to keep record of incomplete exchange processes. A society is said to be in *social equilibrium* if the balances of the exchange values are equilibrated for the successive exchanges occurring along the time.

The analysis of agent social interactions based on the Piaget's theory was first proposed in the various works that led to the paper by Rodrigues and Costa [45] in 2003. Our qualitative approach for the modeling of social exchanges in multiagent systems was introduced in 2005 [16]. In this approach, the exchanges may be performed by the agents according to different observable strategies, which are called the agents' *social exchange strategies*.

In our work [12–15, 38–40], we have been mainly concerned with the problem of the *self-regulation of social exchanges in agent societies*. For that, we have built on hybrid BDI-POMDP agent models, defined over the BDI (Beliefs, Desires, Intentions) architecture [5, 43, 56] with plans derived from POMDP (Partially Observable Markov Decision Processes) [30] models of social exchange strategies, in the line of the works of Simari and Parsons [53] and other authors [33, 34, 37, 51, 52, 54, 55].

However, the *main problem* of the self-regulation of strategy-based social exchanges in *open* agent societies was still to be tackled, namely, "How to deal with the appearance of new social exchange strategies, whenever an agent of the society modifies its strategy (giving rise to different and unexpected social reactions) or a new agent (with a different and unknown strategy) enters the society?"

The present paper[1] advances the solution to the problem of regulating social exchanges in open agent societies, by treating the preliminary problem of *recognizing* and *learning* new models of social exchange strategies.

The aim is to extend the BDI-POMDP model for the self-regulation of social exchanges presented by Pereira et al. [38] with a module based on HMM (Hidden Markov Model) [42] to take into account the problems of recognizing and learning partner agents' social exchange strategies in open societies.

For the recognition problem, the proposed BDI-POMDP-HMM agent proceeds by analyzing the *patterns of refusals* for exchange proposals that are present in a partner agent's behavior.

For the learning problem, the BDI-POMDP-HMM agent uses HMMs to capture the probabilistic state transition and observation functions that model the social exchange strategy of the partner agent, whenever the recognition module was not able to recognize it. The BDI-POMDP-HMM agent then transforms the acquired HMM's probabilistic transition and observation functions into POMDP's probabilistic action-based state transition and observation functions, obtaining a POMDP model of the previously unknown social exchange strategy, thus allowing the extraction of BDI plans for the regulation process, by using the **policyToBDIplans** algorithm [38, 39].

Thus, another challenge that we are addressing in this paper is how to integrate a POMDP model and a HMM, that is, how to obtain a POMDP model from a HMM (and vice versa), since the former has state transition and observation functions based on the actions performed by the agents in each state, whereas in the latter case the state transition and observation functions are not explicitly related to action performances. In this paper, the translation processes from HMM to POMDP and from POMDP to HMM are formally defined, introducing the translation algorithms and some examples.

The paper is organized as follows. Section 2 discusses the modeling of social exchanges in which we have based our works. Section 3 summarizes our work on the regulation of social exchanges in multiagent systems, offering a chronological contextualization and basic concepts necessaries for the development of the paper. In Sect. 4, we present the modeling of the exchanges between social exchange strategy-based agents. The BDI-POMDP model for the self-regulation of social exchange is discussed in Sect. 5. In Sect. 6, we introduce the method for recognizing social exchange strategies. The method for learning new social exchange strategy POMDP models is introduced in Sect. 7, including the discussion on the translation processes from HMM to POMDP and vice versa. Results on simulations of social exchange strategy-based interactions are shown in Sect. 8. Section 9 discusses related work on social exchanges in multiagent systems. Section 10 is the Conclusion.

## 2 Modeling social exchanges

According to Piaget's approach to social interaction [41], a *social exchange* between two agents, $\alpha$ and $\beta$, involves two types of stages.

In stages of type $\mathrm{I}_{\alpha\beta}$, the agent $\alpha$ realizes an action on behalf of (a "service" for) the agent $\beta$, called here a `do-service` action. The *exchange values* involved in this stage are the following:

(i) $r_{\mathrm{I}_{\alpha\beta}}$, which is the value of the *investment* done by $\alpha$ for the realization of a service for $\beta$ (this value is always *negative*);
(ii) $s_{\mathrm{I}_{\beta\alpha}}$, which is the value of $\beta$'s *satisfaction* due to the receiving of the service done by $\alpha$ (this value may be *positive*, *negative*, or *null*);
(iii) $t_{\mathrm{I}_{\beta\alpha}}$, which is the value of $\beta$'s *debt*, the debt it acquired to $\alpha$ for its satisfaction with the service done by $\alpha$ (this value may be *positive*, *negative* or *null*); and
(iv) $v_{\mathrm{I}_{\alpha\beta}}$, which is the value of the *credit* that $\alpha$ acquires from $\beta$ for having realized the service for $\beta$ (this value may be *positive*, *negative* or *null*).

In stages of type $\mathrm{II}_{\alpha\beta}$, the agent $\alpha$ asks the payment for the service previously done for the agent $\beta$, in the form of

[1] This paper is an extended version of the work presented at BWSS 2010—the Second Brazilian Workshop on Social Simulation—and selected by the BWSS 2010 Program Committee as one of the two best works presented at the workshop.

an ask-service action, and the values related with this exchange have similar meaning.

Observe that the order in which the stages may occur is not necessarily $I_{\alpha\beta}$–$II_{\alpha\beta}$.

The values $r_{I_{\alpha\beta}}$, $s_{I_{\beta\alpha}}$, $r_{II_{\beta\alpha}}$ and $s_{II_{\alpha\beta}}$ are called *material values* (investments $r$, and satisfactions $s$), generated by the evaluation of *immediate exchanges*; the values $t_{I_{\beta\alpha}}$, $v_{I_{\alpha\beta}}$, $t_{II_{\beta\alpha}}$ and $v_{II_{\alpha\beta}}$ are the *virtual values* (credits $v$ and debts $t$), concerning exchanges that are expected to happen in the future [41].

The four exchange values involved in an interaction are assigned in ways that are completely uncorrelated to each other, for reasons that may be completely independent of the objective features that characterize the performance of the service (see Example 1, in this section).

A *social exchange process* is composed by a sequence of stages of type $I_{\alpha\beta}$ and/or $II_{\alpha\beta}$, performed in discrete time instants.

The *material results*, according to the points of view of any agents $\alpha$ and $\beta$, are given by the *sum total* [12, 13, 16] of the material values that occurred in the sequence of exchange stages that the agents performed, with the *virtual results* being defined in an analogous way.[2]

A social exchange process is said to be in *material equilibrium* [16] if in all its duration it holds that the pair of material results of $\alpha$ and $\beta$ encloses a given equilibrium point. The *virtual equilibrium* [16] is defined analogously.[3]

Given an on-going interaction, the agents may choose to focus their attention either on the material results or in the virtual results, in order to analyze that interaction.

Material results are important because they report the concrete results obtained from the on-going interaction at each of its steps, and constitute, thus, the main aspect to qualify such interaction.

Virtual results, on the other hand, may be combined with complementary information (like trust and reputation of partner agents) to qualify the possible evolution of the interaction, allowing the agents to make decisions about their ways of participation or non-participation in the future steps of the interaction.

Although it is clear that short-term and long-term aspects of a social interaction are strongly interrelated, we make here a tentative separation between them, in order to simplify our initial study of exchange value-based strategies of social interactions.

So, in our present approach, the material results that agents can obtain in social exchange processes are considered to be the main information that agents should gather in order to distinguish different social exchange strategies to be adopted by the agents concerning the short-term aspects of the interaction. For example, an agent may adopt an exchange strategy that aims to achieve a great amount of material results or, conversely, an exchange strategy that aims to provide a great amount of material results for the other agent (see Sect. 4.1).

On the other hand, the virtual results are considered to be important mainly for decisions concerning the long-term aspects of the interaction. For example, an agent's choice on which agent to offer (require) a service may take into account the debts (credits) the former agent has accumulated in its previous exchanges with each of the other agents of the society. However, the influence of virtual values in the decision on partners for future interactions is not a subject of this paper.

The following example illustrates both that the exchange values need not be correlated to each other and that social exchanges need not result in social equilibrium.

*Example 1* Let agents $A$ and $B$ be interacting in a social context in which there exists a strong social asymmetry between them (e.g., $B$ has a much higher social position then $A$). Let the situation be such that an interaction step arises, with $A$ performing a service for $B$ (on the basis of its social dues to $B$, for instance).

An illustration of the possible exchange values that may arise in such situation is as follows. First, $A$ assigns to its service an investment value. Let us say that such value is denoted by high, meaning for instance that $A$ considers that the service required a high degree of involvement, used a lot of resources, etc.

Second, after the service is finished, $B$ assigns a satisfaction value to it. Let us say that such value is denoted by medium, meaning, for instance, that $B$ considers that the benefit it acquired from $A$'s service ranges among the average benefit $B$ is used to receive from services performed by agents with the same social position as $A$.

Third, $B$ decides upon the debit it has acquired toward $A$ from the performance of the service. Let us say that such value is denoted by low, meaning, for instance, that $B$ considers that not only the satisfaction it got from the service was an average satisfaction but also that $A$ did not care to treat $B$ as adequately as $B$ thought $A$ should have treated it during the performance of the service (say, $A$ did not care to talk to $B$ with a language that properly acknowledged the difference between their social positions).

Finally, $A$ decides upon the credit it has upon $B$ for the service performed. Let us say that such value can be denoted by null, meaning for instance that $A$ considers that it is only an obligation of agents that have social positions similar to its own to perform that kind of service for agents that have social positions similar to that of $B$.

[2] In previous work, we used intervals to represent exchange values, providing a specific interval algebra for their manipulation. See [12, 13, 16] for more details on this formulation.

[3] Notice that Piaget's notion of equilibrium has no game-theoretic meaning, since it involves no notion of game strategy, and concerns just an algebraic sum.

The full exploration of the situation created in Example 1 is out of the scope of this paper, but from it one can hint how an interaction where $A$ charges $B$ for the service performed could restore at least in part the social equilibrium. One can also hint the role that social stratification plays in biasing the agents toward certain evaluations and how such biases can complicate the possibility of achieving equilibrium in social exchanges.

See also [15, 16] for more details on this modeling, and [4, 9, 19, 29, 57] for other approaches related to social exchanges.

## 3 Previous work on the regulation of strategy-based social exchanges in agent societies

The (centralized) mechanism for the regulation of social exchanges in agent societies, based on the concept of *equilibrium supervisor* with an associated Qualitative Interval Markov Decision Process (QI-MDP) was introduced by Dimuro and Costa [12, 13, 16]. That approach was extended by Dimuro et al. [14, 15] to consider *strategy-based social exchanges*.[4] Following, we internalized the regulation mechanism in the agents, through a BDI-POMDP hybrid model [38], going toward the *self-regulation* of social exchanges. A comparison between social exchange regulation models was also presented by Pereira et al. [40]

Observe that *hybrid models* have been proposed in the literature in order to take advantages of both POMDPs and BDI architecture (see, e.g., [33, 34, 37, 51–55]).[5]

Simari and Parsons [53] showed that the plans derived from an optimal policy are exactly the same adopted by a BDI agent that selects plans with the highest utility, and that operate with an optimal reconsideration strategy.

So, for our hybrid BDI-POMDP model [38] to have the social exchange regulation process internalized in the agent model, we introduced the **policyToBDIplans** algorithm, which extracts BDI plans from policy graphs related to optimal policies of POMDP models defined for the different social exchange strategies that the partner agents may follow, one plan for each different exchange strategy [39]. Such plans are said to "obey optimal POMDP policies" (cf., [53]). At each interaction step, one of such plans is put to use, depending on the partner of the interaction and on the current balance of material values.

[4]In our previous works (e.g., [15–17, 38]), the agents's social exchange strategies were called agents' exchange personality traits. However, differently from other works on agent personality traits (e.g., [6–8, 22, 36]), where traits are defined internally to the agent models, our model describes the agents' observable behaviors in terms of state transition and observation functions, so actually modeling agents' observable strategies, hence our change of terminology.

[5]See [1, 31, 35], for other approaches on hybrid agent models in the context of uncertain domains, based on either BDI or MDP models.

In that work, the problem of the decision about the best exchanges that an agent should propose to its partner in order to achieve social equilibrium, or to promote new interactions, was modeled as a global POMDP for each social exchange strategy that its partner is used to adopt. Considering a set of different strategies (e.g., egoistic, altruistic, tolerant strategies), each global POMDP was decomposed into three sub-POMDPs, according to the current internal state (favorable, equilibrated or unfavorable balance of material exchange values) of the agent that was trying to regulate the interaction [38].

For the sake of simplicity, in that work, the agents were not allowed to adopt a strategy different from the ones previously defined for them, which were globally known to the agent society.

That was the problem that motivated the present paper, i.e., the overcoming of the restriction of the recognition process to pre-defined sets of possible exchange strategies, in order to allow for its use in open agent societies.

## 4 Modeling strategy-based social exchanges

In this section, we present the basic concepts of our model of social exchanges between two agents $\alpha$ and $\beta$.

The possible ranges of material results of social exchange processes constitute the internal states of an agent. The *sets of internal states* of the agents $\alpha$ and $\beta$ are then given by

$$\mathrm{E}_\alpha = \{E_\alpha^-, E_\alpha^0, E_\alpha^+\} \quad \text{and} \quad \mathrm{E}_\beta = \{E_\beta^-, E_\beta^0, E_\beta^+\}, \tag{1}$$

respectively, where $E_\alpha^0$ and $E_\beta^0$ denote the range of equilibrated results (e.g., material results around the zero), $E_\alpha^+$ and $E_\beta^+$ represent favorable results (e.g., positive material results), whereas $E_\alpha^-$ and $E_\beta^-$ denote the ranges of unfavorable results (e.g., negative material results). At any time, the current material results are calculated as the *sum total* of the material exchange values, namely, the investment and satisfaction values, as explained in Sect. 2.

By convention, the agent that performs exchange proposals is denoted by $\alpha$, and the agent that has to decide on accepting or refusing $\alpha$'s exchange proposals is represented by $\beta$. The *set of exchange proposals* that the agent $\alpha$ may make to the agent $\beta$ is construed as the set of the two possible actions that $\alpha$ may perform in each exchange stage, given by

$$\mathrm{P} = \{\texttt{do-service}, \texttt{ask-service}\}, \tag{2}$$

with `do-service` meaning that $\alpha$ proposes to perform a service to $\beta$ (exchange stage of type $\mathrm{I}_{\alpha\beta}$), and `ask-service` meaning that $\alpha$ proposes that $\beta$ performs a service to $\alpha$ (exchange stage $\mathrm{II}_{\alpha\beta}$).

**Table 1** Example patterns of operation of the state transition function of a social exchange strategy-based agent $\beta$

(a) Exchange stages of type $\mathrm{I}_{\alpha\beta}$, $\alpha$ performing a `do-service` action

| $\Pi(\mathrm{E}_\beta)$ | Egoistic strategy | | | Altruistic strategy | | | Tolerant strategy | | |
|---|---|---|---|---|---|---|---|---|---|
| | $E^0_\beta$ | $E^+_\beta$ | $E^-_\beta$ | $E^0_\beta$ | $E^+_\beta$ | $E^-_\beta$ | $E^0_\beta$ | $E^+_\beta$ | $E^-_\beta$ |
| $E^0_\beta$ | very-low | very-high | 0.0 | very-high | very-low | 0.0 | very-low | very-high | 0.0 |
| $E^+_\beta$ | 0.0 | 1.0 | 0.0 | 0.0 | 1.0 | 0.0 | 0.0 | 1.0 | 0.0 |
| $E^-_\beta$ | high | high | very-low | low | very-low | very-high | high | high | very-low |

(b) Exchange stages of type $\mathrm{II}_{\alpha\beta}$, $\alpha$ performing an `ask-service` action

| $\Pi(\mathrm{E}_\beta)$ | Egoistic strategy | | | Altruistic strategy | | | Tolerant strategy | | |
|---|---|---|---|---|---|---|---|---|---|
| | $E^0_\beta$ | $E^+_\beta$ | $E^-_\beta$ | $E^0_\beta$ | $E^+_\beta$ | $E^-_\beta$ | $E^0_\beta$ | $E^+_\beta$ | $E^-_\beta$ |
| $E^0_\beta$ | very-high | 0.0 | very-low | very-low | 0.0 | very-high | very-low | 0.0 | very-high |
| $E^+_\beta$ | low | very-high | very-low | high | very-low | high | high | very-low | high |
| $E^-_\beta$ | 0.0 | 0.0 | 1.0 | 0.0 | 0.0 | 1.0 | 0.0 | 0.0 | 1.0 |

### 4.1 The strategy-based social exchange state transition functions

The agents may have different *social exchange strategies* that give rise to different *state transition functions*, defined, for an agent $\beta$, by the function:

$$T : \mathrm{E}_\beta \times \mathrm{P} \rightarrow \Pi(\mathrm{E}_\beta), \tag{3}$$

which specifies, given $\beta$'s current state in $\mathrm{E}_\beta$ and $\alpha$'s exchange proposal in P, a probability distribution over the set of states that $\beta$ will achieve next.

In the following, we illustrate some of those strategies that constitute our *initial set of social exchange strategies*:

**Egoistic Strategy**: the agent is mostly seeking his own benefit, with a very high probability to accept exchanges that represent *transitions to favorable results* (i.e., exchange stages in which the other agent performs a service to it);

**Altruistic Strategy**: the agent is mostly seeking the benefit of the other agent, with a very high probability to accept exchanges that represent transitions toward states where the other agent has favorable results (i.e., exchange stages in which it performs a service to the other agent); in this case, there is a very high probability of *transitions to unfavorable results*;

**Tolerant Strategy**: the agent has a high probability to accept all kinds of exchanges, provided they are sensible from its point of view.

Table 1(a) illustrates the patterns of operation of the probability distribution $\Pi(\mathrm{E}_\beta)$ over the set of internal states $\mathrm{E}_\beta$, as it is determined by the state transition function $T$ that characterizes a social exchange strategy-based agent $\beta$, when another agent $\alpha$ offers to perform a service to $\beta$ (i.e., an exchange stage of type $\mathrm{I}_{\alpha\beta}$, with an action `do-service`).

In Table 1(a), observe that, for an agent $\beta$ with an egoistic strategy, transitions ending in favorable results ($E^+_\beta$) are the most probable, meaning that there is a very high probability that $\beta$ will accept the service proposed by $\alpha$. On the other hand, if $\beta$ is an agent with an altruistic strategy then there is a high probability that it will refuse such proposal, remaining in the same state.

In Table 1(a), the probability of $\beta$ to change this state from $E^0$ to $E^-$ or from $E^+$ to $E^-$ or from $E^+$ to $E^0$, considering any strategy model, is always equals to 0, since there is no possibility of $\beta$ to decrease its material results for receiving a service performed by $\alpha$. However, there is always a probability different from 0 of $\beta$ to remain in the state $E^-$, whenever its current state is $E^-$, which is the case when $\beta$ rejects $\alpha$'s proposal.

See Table 1(b) to compare the converse probabilistic behavior of an agent $\beta$ with an egoistic/altruistic strategy, when $\alpha$ requests a service from $\beta$ (i.e., exchange stage of type $\mathrm{II}_{\alpha\beta}$, with an action `ask-service`). In this table, the probability of $\beta$ to change this state from $E^0$ to $E^+$ or from $E^-$ to $E^+$ or from $E^-$ to $E^0$, in any strategy model, is always equals to 0, since there is no possibility of $\beta$ to increase its material results for performing a service to $\alpha$. On the contrary, there is always a probability of $\beta$ to stay in the state $E^+$, whenever its current state is $E^+$, which is the case when $\beta$ rejects $\alpha$'s proposal.

Observe also that there is a very high probability that an agent $\beta$ with a tolerant strategy will accept any kind of sensible exchange proposal, as shown in the third column of Table 1.

**Table 2** Example patterns of operation of the observation function of a social exchange strategy-based agent $\beta$

(a) Exchange stages of type $\mathrm{I}_{\alpha\beta}$, $\alpha$ performing a `do-service` action

| $\Pi(\Omega)$ | Egoistic strategy | | Altruistic strategy | | Tolerant strategy | |
|---|---|---|---|---|---|---|
| | $A$ | $R$ | $A$ | $R$ | $A$ | $R$ |
| $E^0_\beta$ | high | low | low | high | very-high | very-low |
| $E^+_\beta$ | high | low | very-low | very-high | low | high |
| $E^-_\beta$ | very-high | very-low | low | high | very-high | very-low |

(b) Exchange stages of type $\mathrm{II}_{\alpha\beta}$, $\alpha$ performing an `ask-service` action

| $\Pi(\Omega)$ | Egoistic strategy | | Altruistic strategy | | Tolerant strategy | |
|---|---|---|---|---|---|---|
| | $A$ | $R$ | $A$ | $R$ | $A$ | $R$ |
| $E^0_\beta$ | low | high | high | low | very-high | very-low |
| $E^+_\beta$ | low | high | very-high | very-low | very-high | very-low |
| $E^-_\beta$ | very-low | very-high | high | low | low | high |

### 4.2 The observable social exchange behaviors

Each agent is assumed to have direct access only to its own internal states, that is, the agent can only evaluate its own material results.

However, it is assumed that the agents are able to make observations on each others' exchange behaviors.

The behaviors that an agent $\beta$ manifests for another agent $\alpha$ on $\alpha$'s exchange proposals (e.g., accepting or refusing $\alpha$'s service proposals) constitute the observations that $\alpha$ may make about $\beta$, in order to figure out $\beta$'s social exchange strategy.

So, the *set of observable exchange behaviors* of an agent $\beta$ is given by

$$\Omega = \{A, R\}, \tag{4}$$

where $A$ and $R$ mean that the agent $\beta$ accepts and refuses the exchange proposal, respectively.[6]

An *observation function*, based on the strategy model of an agent $\beta$, in then defined as:

$$O : \mathrm{E}_\beta \times \mathrm{P} \to \Pi(\Omega), \tag{5}$$

which, given $\beta$'s (non-observable) state in $\mathrm{E}_\beta$ and the exchange proposal in P, performed by $\alpha$, gives a probability distribution over the set of possible observable exchange behaviors $\Omega$ (i.e., the probability that $\beta$ accepts or refuses $\alpha$'s exchange proposal).

[6]We remark that $\beta$'s exchange behaviors are considered here by $\alpha$ from an external, observational point of view. That is, we are not dealing with cases where $\alpha$ makes use of psychological theory about personality traits that lead $\beta$ to behave the way it is behaving.

Table 2 illustrates the patterns of operation of the observation function of the exchange behaviors of a social exchange strategy-based agent $\beta$, in each of $\beta$'s possible (non-observable) states, for each of $\alpha$'s possible exchange proposal.

## 5 The BDI-POMDP model for the self-regulation of social exchanges

The self-regulation of social exchanges is performed by hybrid BDI-POMDP agents with an internal social control mechanism based on BDI plans extracted from POMDPs optimal policies, using the algorithm **policyToBDIplans** introduced in previous work [38, 39].

The BDI plans are specified for each known social exchange strategy. Initially, the strategy base is composed only by the social exchange strategies mentioned in Sect. 4.1.

We consider that the exchanges between each pair of agents suffer no influence from the exchanges occurring between other pairs of agents. Also, we assume that the agents do not have access to the internal states of each other, so that an agent's internal decision process has only *partial observability* to the states of the other agents (that is, it has only access to the external aspects of their exchange behaviors).

Thus, the agents are able to observe the social exchange behaviors of the other agents, and so are able to evaluate their *indices of exchange refusals* (see Sect. 6), which allows them to recognize already known social exchange strategies, or strategies similar to those already known.

Once an agent recognizes the social exchange strategy adopted by its partner as equal or similar to one stored in its internal base of exchange strategies (see Sect. 6), it is able to choose a plan for social exchanges, and to propose it to its partner, in order to achieve the equilibrium and/or to promote the continuity of their interaction.

The decision process on the best exchanges an agent $\alpha$ can propose to an agent $\beta$ is modeled as a POMDP, denoted by $\text{POMDP}_{\alpha\beta}$.[7]

For any pair of agents $\alpha$ and $\beta$, the $\text{POMDP}_{\alpha\beta}$ of the decision process internalized in the agent $\alpha$ has its set of states $\mathrm{E}_{\alpha\beta}$ composed by pairs of agent internal states:

$$\mathrm{E}_{\alpha\beta} = \{(E_\alpha^*, E_\beta^\dagger) \mid *, \dagger \in \{-, 0, +\}\}, \tag{6}$$

where $\alpha$'s state $E_\alpha^* \in \mathrm{E}_\alpha$ is known by $\alpha$, but $\beta$'s state $E_\beta^\dagger \in \mathrm{E}_\beta$ is non-observable to it. In the following, we denote each pair $(E_\alpha^*, E_\beta^\dagger)$ by $E_{\alpha\beta}^{*\dagger}$. The set of states of the $\text{POMDP}_{\alpha\beta}$ is then given by

$$\mathrm{E}_{\alpha\beta} = \{E_{\alpha\beta}^{*\dagger} \mid *, \dagger \in \{-, 0, +\}\}. \tag{7}$$

By fixing the internal state $E_\alpha^*$ of the agent $\alpha$ in one of its possible three values in $\mathrm{E}_\alpha = \{E_\alpha^-, E_\alpha^0, E_\alpha^+\}$, a partitioning of the set $\mathrm{E}_{\alpha\beta}$ is possible, obtaining the following sets of states:

$$\mathrm{E}_{\alpha\beta}^- = \{E_{\alpha\beta}^{--}, E_{\alpha\beta}^{-0}, E_{\alpha\beta}^{-+}\}, \tag{8}$$

$$\mathrm{E}_{\alpha\beta}^0 = \{E_{\alpha\beta}^{0-}, E_{\alpha\beta}^{00}, E_{\alpha\beta}^{0+}\}, \tag{9}$$

$$\mathrm{E}_{\alpha\beta}^+ = \{E_{\alpha\beta}^{+-}, E_{\alpha\beta}^{+0}, E_{\alpha\beta}^{++}\}. \tag{10}$$

The partitioning of the set of states of the $\text{POMDP}_{\alpha\beta}$ gives rise to three sub-POMDPs, one for each of $\alpha$'s possible internal state, which we denote by $\text{POMDP}_{\alpha\beta}^-$, $\text{POMDP}_{\alpha\beta}^0$ and $\text{POMDP}_{\alpha\beta}^+$, whenever the current $\alpha$'s state is $E_\alpha^-$, $E_\alpha^0$ or $E_\alpha^+$, respectively.

For simplicity, whenever it is clear from the context, the sets of states $\mathrm{E}_{\alpha\beta}^-$, $\mathrm{E}_{\alpha\beta}^0$ and $\mathrm{E}_{\alpha\beta}^+$, given in (8)–(10), are denoted by $\mathrm{E}_{\alpha\beta}^*$, with $* \in \{0, +, -\}$. Similarly, the sub-POMDPs $\text{POMDP}_{\alpha\beta}^-$, $\text{POMDP}_{\alpha\beta}^0$ and $\text{POMDP}_{\alpha\beta}^+$ are denoted by $\text{POMDP}_{\alpha\beta}^*$.

The partitioning of the full $\text{POMDP}_{\alpha\beta}$ allows to reduce the state space and to obtain directly the BDI plans for each of $\alpha$'s current state. As a consequence, since every exchange stage leads $\alpha$ to change its internal state, $\alpha$ will have to change plans for the next interaction, according to the $\text{POMDP}_{\alpha\beta}^*$ that corresponds to its current state.

Then, for each $* \in \{-, 0, +\}$, we define

[7] Decentralized POMDP's [3] were not considered here, since the agents do not perform actions in parallel; they perform a sequence of alternating individual actions in each exchange stage, with an agent deciding the action to perform after knowing the action previously done by its partner.

**Definition 1** (The $\text{POMDP}_{\alpha\beta}^*$ model of social exchange strategies) The sub-POMDP for $\alpha$'s internalized equilibrium control mechanism when its current state is $E_\alpha^*$, is defined as a tuple of the form:

$$\text{POMDP}_{\alpha\beta}^* = (\mathrm{E}_{\alpha\beta}^*, \mathrm{P}, T^*, \Omega, O^*, R^*), \tag{11}$$

where

(i) $\mathrm{E}_{\alpha\beta}^*$ is the set of $\text{POMDP}_{\alpha\beta}^*$ states, corresponding to the state $*$ of $\alpha$, as given by one of (8)–(10);
(ii) P is the set of exchange proposals available for $\alpha$ to perform in each exchange state, defined in (2);
(iii) $T^* : \mathrm{E}_{\alpha\beta}^* \times \mathrm{P} \to \Pi(\mathrm{E}_{\alpha\beta}^*)$ is the $\text{POMDP}_{\alpha\beta}^*$ state transition function, which embeds the state transition function $T$ of $\beta$'s strategy model given in (3), such that, for all $p \in P$ and $E_\beta^i, E_\beta^j \in \{E_\beta^-, E_\beta^0, E_\beta^+\}$:

$$T^*(E_{\alpha\beta}^{*i}, p)(E_{\alpha\beta}^{*j}) = T(E_\beta^i, p)(E_\beta^j); \tag{12}$$

(iv) $\Omega$ is the set of observations that may be realized by $\alpha$ about $\beta$'s exchange behavior, defined in (4);
(v) $O^* : \mathrm{E}_{\alpha\beta}^* \times \mathrm{P} \to \Pi(\Omega)$ is the observation function, embedding the observation function $O$ of $\beta$'s strategy model given in (5), such that, for all $p \in P$, $E_\beta^i \in \{E_\beta^-, E_\beta^0, E_\beta^+\}$ and $\omega \in \Omega$:

$$O^*(E_{\alpha\beta}^{*i}, p)(\omega) = O(E_\beta^i, p)(w); \tag{13}$$

(vi) $R^* : \mathrm{E}_{\alpha\beta}^* \times \mathrm{P} \to \mathbb{R}$ is the reward function for the agent $\alpha$, giving the expected immediate reward to be gained by $\alpha$ for each exchange effectively performed in each state $E_\beta^i \in \{E_\beta^-, E_\beta^0, E_\beta^+\}$.

The solution of a $\text{POMDP}_{\alpha\beta}^*$, its *optimal policy* [30], in a form of a policy graph, helps the agent $\alpha$ to elaborate exchange proposals that may lead *both agents* toward the equilibrium.

We use the algorithm **policyToBDIplans** to build BDI plans that obey such optimal policies. The BDI plans are represented in the language AgentSpeak of Jason platform [5], the plans being formed by sets of rules of the form:

```
+!State(X1): obs==O -> act(Y), !State(X2).
```

where `X1` represents a belief state concerning the current state of the system (pairs of agents' material results), `O` is the current observation (the acceptation or rejection to the latest exchange proposal) and the rule says that, in such situation, the agent should perform the action `Y` (either an action `do-service` or an action `ask-service`), and change its belief state about the system to `X2`.

For more details on this subject, see our previous work in [38, 39].

## 6 Recognizing social exchange strategies

For the process of recognizing social exchange strategies that an agent may adopt, the *Index of Exchange Refusals* (IER) is defined as the ratio between the number of observed refusals (*NR*) to a kind of exchange proposal (either `ask-service` or `do-service` actions) and the total number of such kind of proposals (*NP*), evaluated over a large enough number of interactions:

$$\mathrm{IER}_{\texttt{ask-service}} = \frac{NR_{\texttt{ask-service}}}{NP_{\texttt{ask-service}}}, \tag{14}$$

$$\mathrm{IER}_{\texttt{do-service}} = \frac{NR_{\texttt{do-service}}}{NP_{\texttt{do-service}}}. \tag{15}$$

The IERs of the initial social exchanges strategies discussed in Sect. 4.1 are globally known by all the agents of the society.

Denote by $\mathbb{I}_{\texttt{ask-service}}$ and $\mathbb{I}_{\texttt{do-service}}$ the sets of all observed $\mathrm{IER}_{\texttt{ask-service}}$ and $\mathrm{IER}_{\texttt{do-service}}$ of a given social exchange strategy, respectively. For each initial social exchange strategies introduced in Sect. 4.1 and each kind of exchange proposal, we compute the related *ranges of IERs*, denoted by RIER and defined as the least interval containing all observed IERs:

$$\mathrm{RIER}_{\texttt{ask-service}} = [\inf \mathbb{I}_{\texttt{ask-service}}, \sup \mathbb{I}_{\texttt{ask-service}}], \tag{16}$$

$$\mathrm{RIER}_{\texttt{do-service}} = [\inf \mathbb{I}_{\texttt{do-service}}, \sup \mathbb{I}_{\texttt{do-service}}], \tag{17}$$

where inf and sup are the supremum and the infimum of a set, respectively.

The ranges of IERs can be used by an agent either to recognize an already known social exchange strategy or to classify some *new social exchange strategy* that its partner may adopt. For that, the agent observes the exchange behavior of its partner for a fixed number of exchange proposals (see Sect. 8), in order to evaluate its IERs. If the IER of this *new strategy* is contained in a previously known range of IERs, then this supposed *new strategy* is considered sufficiently similar to one already known by the agent. Then, it recognizes its partner agent's social exchange strategy as corresponding to that range of IER, and is able to select the adequate plans to equilibrate the interactions with that agent. Otherwise, the agent calculates the range of IERs of this *new strategy* for future comparisons.

Observe that the pattern of refusals is more useful (for the purpose of strategy classification) than the pattern of acceptances. In the former case, the exchange does not occur, and the agents' balances of exchanges remain the same. In the second case, there are a lot of different possibilities of variations on the balances of exchanges.

## 7 Learning social exchange strategies

If an agent is not able to recognize its partner's strategy, then it uses a mechanism based on HMMs in order to discover the state transition and observation functions that best fit the sequence of observations about the partner agent's exchange behavior.

Given two agents $\alpha$ and $\beta$, denote by $\mathrm{HMM}_{\alpha\beta}$ the HMM of the strategy learning mechanism of an agent $\alpha$ that needs to learn the POMDP model of the social exchange strategy adopted by its partner agent $\beta$.

Considering a sequence of observations on $\beta$'s exchange behavior and an arbitrary initial $\mathrm{HMM}_{\alpha\beta}$, it is possible to apply the well-known Baum Welch algorithm [42] in order to optimize the $\mathrm{HMM}_{\alpha\beta}$ for $\beta$'s social exchange strategy. Then, the elements of the optimized $\mathrm{HMM}_{\alpha\beta}$ can be translated into the elements of a $\mathrm{POMDP}_{\alpha\beta}$, allowing the agent $\alpha$ to compute optimal policies for the new social exchange strategy, and, afterwards, to extract the BDI plans for the agent's plan base, together with the respective range of IERs.

In a way similar to the case of the $\mathrm{POMDP}_{\alpha\beta}$ models, the $\mathrm{HMM}_{\alpha\beta}$ models can be partitioned into smaller $\mathrm{HMM}^*_{\alpha\beta}$ models, by restricting the set of state to fixed values of $\alpha$'s states, in order to reduce the state space of the learning procedure.

In the following, we discuss the three main problems of such learning process:

(i) The definition of the $\mathrm{HMM}^*_{\alpha\beta}$ model;
(ii) The choice of the initial values of the $\mathrm{HMM}^*_{\alpha\beta}$ to serve as input data for the Baum Welch algorithm, which imply the translation of an arbitrary $\mathrm{POMDP}^*_{\alpha\beta}$ model into a $\mathrm{HMM}^*_{\alpha\beta}$ model; and
(iii) The translation of the $\mathrm{HMM}^*_{\alpha\beta}$ model that results from the Baum Welch learning algorithm back to a $\mathrm{POMDP}^*_{\alpha\beta}$ model, which consists in the reverse process of (ii).

### 7.1 The HMM model of social exchange strategies

Observe that, although both $\mathrm{HMM}_{\alpha\beta}$ and $\mathrm{POMDP}_{\alpha\beta}$ models have many elements in common, the translation of the $\mathrm{HMM}_{\alpha\beta}$ state transition and observation functions into the *action-dependent* $\mathrm{POMDP}_{\alpha\beta}$ state transition and observation functions cannot be done directly.

In the $\mathrm{POMDP}_{\alpha\beta}$ model, the probability distribution over the set of states is directly linked to the kind of action (either `do-service` or `ask-service` actions) performed by $\alpha$ in each actual state, whereas in the $\mathrm{HMM}_{\alpha\beta}$ those distributions provide overall probabilistic values, independent of the action that might have been performed at each state.

In order to be able to relate the state transition functions of both models, we unify the $\mathrm{POMDP}_{\alpha\beta}$ state transition matrices for both kind of actions into a single extended state transition matrix. For that, the set of states in the $\mathrm{HMM}_{\alpha\beta}$ is extended to specify the kind of action that may be performed by $\alpha$, obtaining:

$$\begin{aligned}\mathrm{EX}_{\alpha\beta} &= \mathrm{E}_\alpha \times \mathrm{E}_\beta \times \mathrm{P}\\ &= \{E^{--}_{\alpha\beta(\mathrm{do})}, E^{-0}_{\alpha\beta(\mathrm{do})}, E^{-+}_{\alpha\beta(\mathrm{do})}, E^{0-}_{\alpha\beta(\mathrm{do})}, E^{00}_{\alpha\beta(\mathrm{do})},\\ &\quad E^{0+}_{\alpha\beta(\mathrm{do})}, E^{+-}_{\alpha\beta(\mathrm{do})}, E^{+0}_{\alpha\beta(\mathrm{do})}, E^{++}_{\alpha\beta(\mathrm{do})}, E^{--}_{\alpha\beta(\mathrm{ask})},\\ &\quad E^{-0}_{\alpha\beta(\mathrm{ask})}, E^{-+}_{\alpha\beta(\mathrm{ask})}, E^{0-}_{\alpha\beta(\mathrm{ask})}, E^{0+}_{\alpha\beta(\mathrm{ask})}, E^{00}_{\alpha\beta(\mathrm{ask})},\\ &\quad E^{+-}_{\alpha\beta(\mathrm{ask})}, E^{+0}_{\alpha\beta(\mathrm{ask})}, E^{++}_{\alpha\beta(\mathrm{ask})}\}.\end{aligned} \tag{18}$$

With such set of states, the restriction to particular values of states of $\alpha$ in $\mathrm{E}_\alpha = \{E^-_\alpha, E^0_\alpha, E^+_\alpha\}$ amounts to define the following sets of states:

$$\begin{aligned}\mathrm{EX}^-_{\alpha\beta} &= \{E^-_\alpha\} \times \mathrm{E}_\beta \times \mathrm{P}\\ &= \{E^{--}_{\alpha\beta(\mathrm{do})}, E^{-0}_{\alpha\beta(\mathrm{do})}, E^{-+}_{\alpha\beta(\mathrm{do})}, E^{--}_{\alpha\beta(\mathrm{ask})},\\ &\quad E^{-0}_{\alpha\beta(\mathrm{ask})}, E^{-+}_{\alpha\beta(\mathrm{ask})}\}.\end{aligned} \tag{19}$$

$$\begin{aligned}\mathrm{EX}^0_{\alpha\beta} &= \{E^0_\alpha\} \times \mathrm{E}_\beta \times \mathrm{P}\\ &= \{E^{0-}_{\alpha\beta(\mathrm{do})}, E^{00}_{\alpha\beta(\mathrm{do})}, E^{0+}_{\alpha\beta(\mathrm{do})}, E^{0-}_{\alpha\beta(\mathrm{ask})},\\ &\quad E^{00}_{\alpha\beta(\mathrm{ask})}, E^{0+}_{\alpha\beta(\mathrm{ask})}\}.\end{aligned} \tag{20}$$

$$\begin{aligned}\mathrm{EX}^+_{\alpha\beta} &= \{E^+_\alpha\} \times \mathrm{E}_\beta \times \mathrm{P}\\ &= \{E^{+-}_{\alpha\beta(\mathrm{do})}, E^{+0}_{\alpha\beta(\mathrm{do})}, E^{++}_{\alpha\beta(\mathrm{do})}, E^{+-}_{\alpha\beta(\mathrm{ask})},\\ &\quad E^{+0}_{\alpha\beta(\mathrm{ask})}, E^{++}_{\alpha\beta(\mathrm{ask})}\}.\end{aligned} \tag{21}$$

Analogously to what was obtained in Sect. 5, the partitioning of the set of states of the $\mathrm{HMM}_{\alpha\beta}$ gives rise to three sub-HMMs, one for each of $\alpha$'s possible internal state, which we denote by $\mathrm{HMM}^-_{\alpha\beta}$, $\mathrm{HMM}^0_{\alpha\beta}$ and $\mathrm{HMM}^+_{\alpha\beta}$, whenever the current $\alpha$'s state is $E^-_\alpha$, $E^0_\alpha$ or $E^+_\alpha$, respectively.

Again, for simplicity, whenever it is clear from the context, the sets of states $\mathrm{EX}^-_{\alpha\beta}$, $\mathrm{EX}^0_{\alpha\beta}$ and $\mathrm{EX}^+_{\alpha\beta}$, given in (19)–(21), are denoted by $\mathrm{EX}^*_{\alpha\beta}$, with $* \in \{0, +, -\}$. Similarly, the sub-HMMs $\mathrm{HMM}^-_{\alpha\beta}$, $\mathrm{HMM}^0_{\alpha\beta}$ and $\mathrm{HMM}^+_{\alpha\beta}$ are denoted by $\mathrm{HMM}^*_{\alpha\beta}$.

Then, for each $* \in \{-, 0, +\}$, we define

**Definition 2** (The $\mathrm{HMM}^*_{\alpha\beta}$ of social exchange strategies) The sub-HMM for $\alpha$'s learning mechanism when its current state is $E^*_\alpha$, is defined as a tuple of the form:

$$\mathrm{HMM}^*_{\alpha\beta} = \left(\mathrm{EX}^*_{\alpha\beta}, \Pi^0_{\mathrm{EX}^*_{\alpha\beta}}, TX^*, \Omega, OX^*\right), \tag{22}$$

where

(i) $\mathrm{EX}^*_{\alpha\beta}$ is the extended set of states corresponding to $\alpha$'s state $E^*_\alpha$, given by one of (19)–(21);
(ii) $\Pi^0_{\mathrm{EX}^*_{\alpha\beta}}$ is the initial probability distribution of the extended set of states;
(iii) $TX^* : \mathrm{EX}^*_{\alpha\beta} \to \Pi(\mathrm{EX}^*_{\alpha\beta})$ is the state transition function patterned on $\beta$'s strategy model, which, given the current state in $\mathrm{EX}^*_{\alpha\beta}$, gives a probability distribution over the set $\mathrm{EX}^*_{\alpha\beta}$ of states;
(iv) $\Omega$ is the set of observations that may be realized by $\alpha$ about $\beta$'s exchange behavior (4);
(v) $OX^* : \mathrm{EX}^*_{\alpha\beta} \to \Pi(\Omega)$ is the observation function patterned on $\beta$'s strategy model, which, given the current state in $\mathrm{EX}^*_{\alpha\beta}$, indicates a probability distribution over the set $\Omega$ of possible observations.

Observe that there is an isomorphism between the set $\mathrm{EX}^*_{\alpha\beta}$, the domain of the state transition and observation functions of a $\mathrm{HMM}^*_{\alpha\beta}$ (respectively, $TX^*$ and $OX^*$, in Definition 2), and the set $\mathrm{E}^*_{\alpha\beta} \times \mathrm{P}$, the domain of the state transition and observation functions of a $\mathrm{POMDP}^*_{\alpha\beta}$ (respectively, $T^*$ and $O^*$, in Definition 1). Such isomorphism, defined, for all $p \in \mathrm{P} = \{\texttt{do-service}, \texttt{ask-service}\}$ and $*, \dagger \in \{-, +, 0\}$, by

$$\Psi : \mathrm{EX}^*_{\alpha\beta} \to \mathrm{E}^*_{\alpha\beta} \times \mathrm{P}, \tag{23}$$

such that

$$\Psi\left(E^{*\dagger}_{\alpha\beta(p)}\right) = \left(E^{*\dagger}_{\alpha\beta}, p\right), \tag{24}$$

guarantees that the translation processes between the models, in both directions, is viable, as explained in the following sections.

### 7.2 Preparing the input HMM for the Baum Welch Algorithm: the translation process from POMDPs to HMMs

The elements of the input $\mathrm{HMM}^*_{\alpha\beta}$ for the Baum Welch Algorithm may be defined arbitrarily, or may be based in the known $\mathrm{POMDP}^*_{\alpha\beta}$ strategy model that has the most similar IERs to the IERs of the unknown strategy, which, however, is not contained in any known RIER.

The translation process of the state transition function of such $\mathrm{POMDP}^*_{\alpha\beta}$ into the state transition function of a $\mathrm{HMM}^*_{\alpha\beta}$ is performed by the algorithm $\mathbf{T_{POMDP}toT_{HMM}}$ (Algorithm 1).

Considering that the agent $\alpha$ performs, during the learning process, the action `do-service` with *a priori* probability $\pi_{(\mathrm{do})}$ and the action `ask-service` with *a priori*

**input**: $T_{(\text{do})}, T_{(\text{ask})}$: $3 \times 3$ matrices of probability values, representing the $\text{POMDP}^*_{\alpha\beta}$ state transition function $T^* : \text{E}^*_{\alpha\beta} \rightarrow \Pi(\text{E}^*_{\alpha\beta})$, for each kind of exchange proposal; $\pi_{(\text{do})}, \pi_{(\text{ask})}$: the probabilities of each kind of exchange proposal;
**output**: $TX$: a $3 \times 3$ matrix of probability values, representing the $\text{HMM}^*_{\alpha\beta}$ state transition function $TX^* : \text{EX}^*_{\alpha\beta} \rightarrow \Pi(\text{EX}^*_{\alpha\beta})$;

{ *For any matrix* $M = [m_{EE'}]$, *the expression* $M[E, E']$ *denotes the entry* $m_{EE'}$ }

```
begin
  for i ∈ {−, 0, +} do
    for j ∈ {−, 0, +} do
      for k1, k2 ∈ {do, ask} do
        TX[E*i_αβ(k1), E*j_αβ(k2)] ⟵
        π(k2) · T(k1)[E*i_αβ, E*j_αβ];
      end
    end
  end
end
```

**Algorithm 1**: Algorithm $\mathbf{T_{POMDP}toT_{HMM}}$ for translating $\text{POMDP}^*_{\alpha\beta}$ transition functions into $\text{HMM}^*_{\alpha\beta}$ transition functions

probability $\pi_{(\text{ask})}$, such that $\pi_{(\text{do})} + \pi_{(\text{ask})} = 1$, then this algorithm obtains the $\text{HMM}^*_{\alpha\beta}$ transition function $TX^*$ defined by

$$TX^* : \text{EX}^*_{\alpha\beta} \rightarrow \Pi\big(\text{EX}^*_{\alpha\beta}\big), \tag{25}$$

where, for each state $E^{*i}_{\alpha\beta(\text{do})}, E^{*j}_{\alpha\beta(\text{do})}, E^{*i}_{\alpha\beta(\text{ask})}, E^{*j}_{\alpha\beta(\text{ask})} \in \text{EX}^*_{\alpha\beta}$

$$TX^*\big(E^{*i}_{\alpha\beta(\text{do})}\big)\big(E^{*j}_{\alpha\beta(\text{do})}\big) = \pi_{(\text{do})} \cdot T^*(\texttt{do-service})\big(E^{*i}_{\alpha\beta}\big)\big(E^{*j}_{\alpha\beta}\big),$$

$$TX^*\big(E^{*i}_{\alpha\beta(\text{do})}\big)\big(E^{*j}_{\alpha\beta(\text{ask})}\big) = \pi_{(\text{ask})} \cdot T^*(\texttt{do-service})\big(E^{*i}_{\alpha\beta}\big)\big(E^{*j}_{\alpha\beta}\big),$$

$$TX^*\big(E^{*i}_{\alpha\beta(\text{ask})}\big)\big(E^{*j}_{\alpha\beta(\text{do})}\big) = \pi_{(\text{do})} \cdot T^*(\texttt{ask-service})\big(E^{*i}_{\alpha\beta}\big)\big(E^{*j}_{\alpha\beta}\big),$$

$$TX^*\big(E^{*i}_{\alpha\beta(\text{ask})}\big)\big(E^{*j}_{\alpha\beta(\text{ask})}\big) = \pi_{(\text{ask})} \cdot T^*(\texttt{ask-service})\big(E^{*i}_{\alpha\beta}\big)\big(E^{*j}_{\alpha\beta}\big)$$

**input**: $O_{(\text{do})}, O_{(\text{ask})}$: $3 \times 2$ matrices of probability values, representing the $\text{POMDP}^*_{\alpha\beta}$ observation function $O^* : \text{E}^*_{\alpha\beta} \rightarrow \Pi(\Omega)$, for each kind of exchange proposal
**output**: $OX$: a $6 \times 2$ matrix of probability values, representing the $\text{HMM}^*_{\alpha\beta}$ observation function $OX^* : \text{EX}^*_{\alpha\beta} \rightarrow \Pi(\Omega)$

{ *For any matrix* $M = [m_{EE'}]$, *the expression* $M[E]$ *denotes the row* $E$ *of* $M$ }

```
begin
  for i ∈ {−, 0, +} do
    OX[E*i_αβ(do)] ⟵ O(do)[E*i_αβ];
    OX[E*i_αβ(ask)] ⟵ O(ask)[E*i_αβ];
  end
end
```

**Algorithm 2**: Algorithm $\mathbf{O_{POMDP}toO_{HMM}}$ for translating $\text{POMDP}^*_{\alpha\beta}$ observation functions into $\text{HMM}^*_{\alpha\beta}$ observation functions

with $i, j \in \{-, 0, +\}$, $\texttt{do-service}, \texttt{ask-service} \in \text{P}$, $E^{*i}_{\alpha\beta}, E^{*j}_{\alpha\beta} \in \text{E}^*_{\alpha\beta}$, and $\text{E}^*_{\alpha\beta}, T^*, \text{P}$ as specified in Definition 1, and $\text{EX}^*_{\alpha\beta}, TX^*$ as specified in Definition 2.

The translation process of the $\text{POMDP}^*_{\alpha\beta}$ observation function into a $\text{HMM}^*_{\alpha\beta}$ observation function is performed by the algorithm $\mathbf{O_{POMDP}toO_{HMM}}$ (Algorithm 2). This algorithm gives the $\text{HMM}^*_{\alpha\beta}$ observation function $OX^*$ defined by

$$OX^* : \text{EX}^*_{\alpha\beta} \rightarrow \Pi(\Omega), \tag{26}$$

such that, for each state $E^{*i}_{\alpha\beta(\text{do})}, E^{*i}_{\alpha\beta(\text{ask})} \in \text{EX}^*_{\alpha\beta}$, and observation $\omega \in \Omega$:

$$OX^*\big(E^{*i}_{\alpha\beta(\text{do})}\big)(\omega) = O^*(\texttt{do-service})\big(E^{*i}_{\alpha\beta}\big)(\omega),$$

$$OX^*\big(E^{*i}_{\alpha\beta(\text{ask})}\big)(\omega) = O^*(\texttt{ask-service})\big(E^{*i}_{\alpha\beta}\big)(\omega)$$

with $i \in \{-, 0, +\}$, $\texttt{do-service}, \texttt{ask-service} \in \text{P}$, $E^{*i}_{\alpha\beta} \in \text{E}^*_{\alpha\beta}$, and $\Omega, \text{E}^*_{\alpha\beta}, O^*, \text{P}$ as specified in Definition 1, and $\text{EX}^*_{\alpha\beta}, OX^*$ as specified in Definition 2.

*Example 2* Suppose that the IERs presented by a social exchange strategy adopted by the agent $\beta$ is not contained in any range of IERs known by the agent $\alpha$. Assume that the strategy model that presents the most similar IERs to $\beta$'s exchange behavior identifies a $\text{POMDP}^*_{\alpha\beta}$ model, whose state transition function $T^*$ and observation function $O^*$, for a given $*$, are given in Table 3 and Table 4, respectively.

Consider that the agent $\alpha$, for the learning process, uses a uniform probability distribution over the set P of exchanges proposals, that is, $\pi_{(\text{do})} = \pi_{(\text{ask})} = 0.5$. Then, the applications of the Algorithms 1 and 2 produce the $\text{HMM}^*_{\alpha\beta}$ state transition function $TX^*$ and observation function $OX^*$ presented in Table 5 and Table 6, respectively.

**Table 3** The $\text{POMDP}^*_{\alpha\beta}$ state transition function $T^*$ of Example 2

(a) Exchange stages of type $\text{I}_{\alpha\beta}$, with $\alpha$ performing a do-service action

| $\Pi(\text{E}^*_{\alpha\beta})$ | $E^{*0}_{\alpha\beta}$ | $E^{*+}_{\alpha\beta}$ | $E^{*-}_{\alpha\beta}$ |
|---|---|---|---|
| $E^{*0}_{\alpha\beta}$ | 0.80 | 0.20 | 0.00 |
| $E^{*+}_{\alpha\beta}$ | 0.00 | 1.00 | 0.00 |
| $E^{*-}_{\alpha\beta}$ | 0.20 | 0.20 | 0.60 |

(b) Exchange stages of type $\text{II}_{\alpha\beta}$, with $\alpha$ performing an ask-service action

| $\Pi(\text{E}^*_{\alpha\beta})$ | $E^{*0}_{\alpha\beta}$ | $E^{*+}_{\alpha\beta}$ | $E^{*-}_{\alpha\beta}$ |
|---|---|---|---|
| $E^{*0}_{\alpha\beta}$ | 0.80 | 0.00 | 0.20 |
| $E^{*+}_{\alpha\beta}$ | 0.20 | 0.60 | 0.20 |
| $E^{*-}_{\alpha\beta}$ | 0.00 | 0.00 | 1.00 |

**Table 4** The $\text{POMDP}^*_{\alpha\beta}$ observation function $O^*$ of Example 2

(a) Exchange stages of type $\text{I}_{\alpha\beta}$, with $\alpha$ performing a do-service action

| $\Pi(\Omega)$ | $A$ | $R$ |
|---|---|---|
| $E^{*0}_{\alpha\beta}$ | 0.50 | 0.50 |
| $E^{*+}_{\alpha\beta}$ | 0.30 | 0.70 |
| $E^{*-}_{\alpha\beta}$ | 0.35 | 0.65 |

(b) Exchange stages of type $\text{II}_{\alpha\beta}$, with $\alpha$ performing an ask-service action

| $\Pi(\Omega)$ | $A$ | $R$ |
|---|---|---|
| $E^{*0}_{\alpha\beta}$ | 0.20 | 0.80 |
| $E^{*+}_{\alpha\beta}$ | 0.30 | 0.70 |
| $E^{*-}_{\alpha\beta}$ | 0.15 | 0.85 |

Finally, considering that the exchange process for the learning process starts in equilibrium, the initial probability distribution $\Pi^0_{\text{EX}_{\alpha\beta}}$ of the extended set of states (Definition 2(ii)) is given in Table 7. $\Pi^0_{\text{EX}_{\alpha\beta}}$, together with the functions $TX^*$ (Table 5) and $OX^*$ (Table 6), constitute the input data for the Baum Welch algorithm.

### 7.3 Obtaining a new social exchange strategy model: the translation process from HHMs to POMDPs

Once the BDI-POMDP-HMM agent $\alpha$ has its $\text{HMM}^*_{\alpha\beta}$ optimized by using the Baum Welch algorithm, then, by a reverse process, it can compress the state space $\text{EX}^*_{\alpha\beta}$ into the original one $\text{E}^*_{\alpha\beta}$, making explicit the implicit action argument of the $\text{HMM}^*_{\alpha\beta}$ one-place transition and observation functions, and then generating the two-place $\text{POMDP}^*_{\alpha\beta}$ transition and observation functions.

The translation process of the $\text{HMM}^*_{\alpha\beta}$ state transition function into the $\text{POMDP}^*_{\alpha\beta}$ state transition function is performed by the algorithm $\mathbf{T_{HMM}toT_{POMDP}}$ (Algorithm 3). This algorithm gives the $\text{POMDP}^*_{\alpha\beta}$ transition function $T^*$ defined by two transition functions, one for each kind of exchange proposal in $\text{P} = \{\texttt{do-service}, \texttt{ask-service}\}$, which are given by

$$T^*(\texttt{do-service}) : \text{E}^*_{\alpha\beta} \to \Pi\big(\text{E}^*_{\alpha\beta}\big), \tag{27}$$

$$T^*(\texttt{ask-service}) : \text{E}^*_{\alpha\beta} \to \Pi\big(\text{E}^*_{\alpha\beta}\big), \tag{28}$$

where, for each state $E^{*i}_{\alpha\beta}, E^{*j}_{\alpha\beta} \in \text{E}^*_{\alpha\beta}$:

$$T^*(\texttt{do-service})\big(E^{*i}_{\alpha\beta}\big)\big(E^{*j}_{\alpha\beta}\big) = TX^*\big(E^{*i}_{\alpha\beta(\text{do})}\big)\big(E^{*j}_{\alpha\beta(\text{do})}\big) + TX^*\big(E^{*i}_{\alpha\beta(\text{do})}\big)\big(E^{*j}_{\alpha\beta(\text{ask})}\big)$$

and

$$T^*(\texttt{ask-service})\big(E^{*i}_{\alpha\beta}\big)\big(E^{*j}_{\alpha\beta}\big) = TX^*\big(E^{*i}_{\alpha\beta(\text{ask})}\big)\big(E^{*j}_{\alpha\beta(\text{do})}\big) + TX^*\big(E^{*i}_{\alpha\beta(\text{ask})}\big)\big(E^{*j}_{\alpha\beta(\text{ask})}\big),$$

**Table 5** The $\text{HMM}^*_{\alpha\beta}$ state transition function $TX^*$ of Example 2

| $\Pi(\text{EX}^*_{\alpha\beta})$ | $E^{*0}_{\alpha\beta(\text{do})}$ | $E^{*0}_{\alpha\beta(\text{ask})}$ | $E^{*+}_{\alpha\beta(\text{do})}$ | $E^{*+}_{\alpha\beta(\text{ask})}$ | $E^{*-}_{\alpha\beta(\text{do})}$ | $E^{*-}_{\alpha\beta(\text{ask})}$ |
|---|---|---|---|---|---|---|
| $E^{*0}_{\alpha\beta(\text{do})}$ | 0.40 | 0.40 | 0.10 | 0.10 | 0.00 | 0.00 |
| $E^{*0}_{\alpha\beta(\text{ask})}$ | 0.40 | 0.40 | 0.00 | 0.00 | 0.10 | 0.10 |
| $E^{*+}_{\alpha\beta(\text{do})}$ | 0.00 | 0.00 | 0.50 | 0.50 | 0.00 | 0.00 |
| $E^{*+}_{\alpha\beta(\text{ask})}$ | 0.10 | 0.10 | 0.30 | 0.30 | 0.10 | 0.10 |
| $E^{*-}_{\alpha\beta(\text{do})}$ | 0.10 | 0.10 | 0.10 | 0.10 | 0.30 | 0.30 |
| $E^{*-}_{\alpha\beta(\text{ask})}$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.50 | 0.50 |

**Table 6** The $\mathrm{HMM}^*_{\alpha\beta}$ observation function $OX^*$ of Example 2

| $\Pi(\Omega)$ | $A$ | $R$ |
|---|---|---|
| $E^{*0}_{\alpha\beta(\mathrm{do})}$ | 0.50 | 0.50 |
| $E^{*0}_{\alpha\beta(\mathrm{ask})}$ | 0.20 | 0.80 |
| $E^{*+}_{\alpha\beta(\mathrm{do})}$ | 0.30 | 0.70 |
| $E^{*+}_{\alpha\beta(\mathrm{ask})}$ | 0.30 | 0.70 |
| $E^{*-}_{\alpha\beta(\mathrm{do})}$ | 0.25 | 0.65 |
| $E^{*-}_{\alpha\beta(\mathrm{ask})}$ | 0.15 | 0.85 |

**Table 7** The initial probability distribution $\Pi^0_{\mathrm{EX}^*_{\alpha\beta}}$

| State | Probability |
|---|---|
| $E^{*0}_{\alpha\beta(\mathrm{do})}$ | 0.50 |
| $E^{*0}_{\alpha\beta(\mathrm{ask})}$ | 0.50 |
| $E^{*+}_{\alpha\beta(\mathrm{do})}$ | 0.00 |
| $E^{*+}_{\alpha\beta(\mathrm{ask})}$ | 0.00 |
| $E^{*-}_{\alpha\beta(\mathrm{do})}$ | 0.00 |
| $E^{*-}_{\alpha\beta(\mathrm{ask})}$ | 0.00 |

with $i, j \in \{0, +, -\}$, $E^{*i}_{\alpha\beta(\mathrm{do})}, E^{*j}_{\alpha\beta(\mathrm{do})}, E^{*i}_{\alpha\beta(\mathrm{ask})}, E^{*j}_{\alpha\beta(\mathrm{ask})} \in \mathrm{EX}^*_{\alpha\beta}$, and $\mathrm{E}^*_{\alpha\beta}$, $T^*$ as specified in Definition 1, and $\mathrm{EX}^*_{\alpha\beta}$, $TX^*$ as specified in Definition 2.

The translation process of the $\mathrm{HMM}^*_{\alpha\beta}$ observation function into the $\mathrm{POMDP}^*_{\alpha\beta}$ observation function is performed by the algorithm $\mathbf{O_{HMM}toO_{POMDP}}$ (Algorithm 4). This algorithm returns the $\mathrm{POMDP}^*_{\alpha\beta}$ observation function $O^*$ defined by two observation functions, one for each kind of exchange proposal in $\mathrm{P} = \{\texttt{do-service}, \texttt{ask-service}\}$, which are given by

$$O^*(\texttt{do-service}) : \mathrm{E}^*_{\alpha\beta} \to \Pi(\Omega), \tag{29}$$

$$O^*(\texttt{ask-service}) : \mathrm{E}^*_{\alpha\beta} \to \Pi(\Omega), \tag{30}$$

such that, for each state $E^{*i}_{\alpha\beta} \in \mathrm{E}^*_{\alpha\beta}$, and observation $\omega \in \Omega$:

$$O^*(\texttt{do-service})\big(E^{*i}_{\alpha\beta}\big)(\omega) = OX^*\big(E^{*i}_{\alpha\beta(\mathrm{do})}\big)(\omega),$$

$$O^*(\texttt{ask-service})\big(E^{*i}_{\alpha\beta}\big)(\omega) = OX^*\big(E^{*i}_{\alpha\beta(\mathrm{ask})}\big)(\omega),$$

with $i \in \{0, +, -\}$, $E^{*i}_{\alpha\beta(\mathrm{do})}, E^{*i}_{\alpha\beta(\mathrm{ask})} \in \mathrm{EX}^*_{\alpha\beta}$, and $\Omega, \mathrm{E}^*_{\alpha\beta}$, $O^*$ as specified in Definition 1, and $\mathrm{EX}^*_{\alpha\beta}$, $OX^*$ as specified in Definition 2.

*Example 3* Considering the input $\mathrm{HMM}^*_{\alpha\beta}$ constructed in Example 2, and a sequence of observations done by $\alpha$ when

**input**: $TX$: a $6 \times 6$ matrix of probability values, representing the $\mathrm{HMM}^*_{\alpha\beta}$ state transition function $TX^* : \mathrm{EX}^*_{\alpha\beta} \to \Pi(\mathrm{EX}^*_{\alpha\beta})$;
**output**: $T_{(\mathrm{do})}, T_{(\mathrm{ask})}$: $3 \times 3$ matrices of probability values, representing the $\mathrm{POMDP}^*_{\alpha\beta}$ state transition function $T^* : \mathrm{E}^*_{\alpha\beta} \to \Pi(\mathrm{E}^*_{\alpha\beta})$, for each kind of exchange proposal;

{ *For any matrix* $M = [m_{EE'}]$, *the expression* $M[E, E']$ *denotes the entry* $m_{EE'}$ }

**begin**
  **for** $i \in \{-, 0, +\}$ **do**
    **for** $j \in \{-, 0, +\}$ **do**
      $T_{(\mathrm{do})}[E^{*i}_{\alpha\beta}, E^{*j}_{\alpha\beta}] \longleftarrow TX[E^{*i}_{\alpha\beta(\mathrm{do})}, E^{*j}_{\alpha\beta(\mathrm{do})}] + T[E^{*i}_{\alpha\beta(\mathrm{do})}, E^{*j}_{\alpha\beta(\mathrm{ask})}]$;
      $T_{(\mathrm{ask})}[E^{*i}_{\alpha\beta}, E^{*j}_{\alpha\beta}] \longleftarrow TX[E^{*i}_{\alpha\beta(\mathrm{ask})}, E^{*j}_{\alpha\beta(\mathrm{do})}] + T[E^{*i}_{\alpha\beta(\mathrm{ask})}, E^{*j}_{\alpha\beta(\mathrm{ask})}]$;
    **end**
  **end**
**end**

**Algorithm 3**: Algorithm $\mathbf{T_{HMM}toT_{POMDP}}$ for translating $\mathrm{HMM}^*_{\alpha\beta}$ transition functions into $\mathrm{POMDP}^*_{\alpha\beta}$ transition functions

**input**: $OX$: a $6 \times 2$ matrix of probability values, representing the $\mathrm{HMM}^*_{\alpha\beta}$ observation function $OX^* : \mathrm{EX}^* \to \Pi(\Omega)$;
**output**: $O_{(\mathrm{do})}, O_{(\mathrm{ask})}$: $3 \times 2$ matrices of probability values, representing the $\mathrm{POMDP}^*_{\alpha\beta}$ observation function $O^* : \mathrm{E}^*_{\alpha\beta} \to \Pi(\Omega)$, for each kind of exchange proposal;

{*For any matrix* $M = [m_{EE'}]$, *the expression* $M[E]$ *denotes the row* $E$ *of* $M$ }

**begin**
  **for** $i \in \{-, 0, +\}$ **do**
    $O_{\alpha\beta(\mathrm{do})}[E^{*i}_{\alpha\beta}] \longleftarrow OX[E^{*i}_{\alpha\beta(\mathrm{do})}]$;
    $O_{\alpha\beta(\mathrm{ask})}[E^{*i}_{\alpha\beta}] \longleftarrow OX[E^{*i}_{\alpha\beta(\mathrm{ask})}]$;
  **end**
**end**

**Algorithm 4**: Algorithm $\mathbf{O_{HMM}toO_{POMDP}}$ for translating $\mathrm{HMM}^*_{\alpha\beta}$ observation functions into $\mathrm{POMDP}^*_{\alpha\beta}$ observation functions

**Table 8** The optimized $HMM^*_{\alpha\beta}$ state transition function $TX^*$ of Example 3

| $\Pi(EX^*_{\alpha\beta})$ | $E^{*0}_{\alpha\beta(do)}$ | $E^{*0}_{\alpha\beta(ask)}$ | $E^{*+}_{\alpha\beta(do)}$ | $E^{*+}_{\alpha\beta(ask)}$ | $E^{*-}_{\alpha\beta(do)}$ | $E^{*-}_{\alpha\beta(ask)}$ |
|---|---|---|---|---|---|---|
| $E^{*0}_{\alpha\beta(do)}$ | 0.52 | 0.16 | 0.16 | 0.16 | 0.00 | 0.00 |
| $E^{*0}_{\alpha\beta(ask)}$ | 0.49 | 0.43 | 0.00 | 0.00 | 0.04 | 0.04 |
| $E^{*+}_{\alpha\beta(do)}$ | 0.00 | 0.00 | 0.54 | 0.46 | 0.00 | 0.00 |
| $E^{*+}_{\alpha\beta(ask)}$ | 0.08 | 0.23 | 0.34 | 0.35 | 0.00 | 0.00 |
| $E^{*-}_{\alpha\beta(do)}$ | 0.10 | 0.14 | 0.0 | 0.0 | 0.39 | 0.38 |
| $E^{*-}_{\alpha\beta(ask)}$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.50 | 0.50 |

**Table 9** The optimized $HMM^*_{\alpha\beta}$ observation function $OX^*$ of Example 3

| $\Pi(\Omega)$ | $A$ | $R$ |
|---|---|---|
| $E^{*0}_{\alpha\beta(do)}$ | 0.56 | 0.44 |
| $E^{*0}_{\alpha\beta(ask)}$ | 0.02 | 0.98 |
| $E^{*+}_{\alpha\beta(do)}$ | 0.18 | 0.82 |
| $E^{*+}_{\alpha\beta(ask)}$ | 0.21 | 0.79 |
| $E^{*-}_{\alpha\beta(do)}$ | 0.21 | 0.79 |
| $E^{*-}_{\alpha\beta(ask)}$ | 0.27 | 0.73 |

**Table 10** The new $POMDP^*_{\alpha\beta}$ state transition function $T^*$ of Example 3

(a) Exchange stages of type $I_{\alpha\beta}$, with $\alpha$ performing a `do-service` action

| $\Pi(E^*_{\alpha\beta})$ | $E^{*0}_{\alpha\beta}$ | $E^{*+}_{\alpha\beta}$ | $E^{*-}_{\alpha\beta}$ |
|---|---|---|---|
| $E^{*0}_{\alpha\beta}$ | 0.68 | 0.32 | 0.00 |
| $E^{*+}_{\alpha\beta}$ | 0.00 | 1.00 | 0.00 |
| $E^{*-}_{\alpha\beta}$ | 0.24 | 0.00 | 0.76 |

(b) Exchange stages of type $II_{\alpha\beta}$, with $\alpha$ performing an `ask-service` action

| $\Pi(E^*_{\alpha\beta})$ | $E^{*0}_{\alpha\beta}$ | $E^{*+}_{\alpha\beta}$ | $E^{*-}_{\alpha\beta}$ |
|---|---|---|---|
| $E^{*0}_{\alpha\beta}$ | 0.92 | 0.00 | 0.08 |
| $E^{*+}_{\alpha\beta}$ | 0.31 | 0.69 | 0.20 |
| $E^{*-}_{\alpha\beta}$ | 0.00 | 0.00 | 1.00 |

**Table 11** The new $POMDP^*_{\alpha\beta}$ observation function $O^*$ of Example 3

(a) Exchange stages of type $I_{\alpha\beta}$, with $\alpha$ performing a `do-service` action

| $\Pi(\Omega)$ | $A$ | $R$ |
|---|---|---|
| $E^{*0}_{\alpha\beta}$ | 0.56 | 0.44 |
| $E^{*+}_{\alpha\beta}$ | 0.18 | 0.82 |
| $E^{*-}_{\alpha\beta}$ | 0.21 | 0.79 |

(b) Exchange stages of type $II_{\alpha\beta}$, with $\alpha$ performing an `ask-service` action

| $\Pi(\Omega)$ | $A$ | $R$ |
|---|---|---|
| $E^{*0}_{\alpha\beta}$ | 0.02 | 0.98 |
| $E^{*+}_{\alpha\beta}$ | 0.21 | 0.79 |
| $E^{*-}_{\alpha\beta}$ | 0.27 | 0.73 |

performing exchange proposals randomly, the application of the Baum Welch algorithm produces the optimized $HMM^*_{\alpha\beta}$ specified by the state transition and observation functions given in Table 8 and Table 9, respectively.

Then, the applications of the Algorithms 3 and 4 produce the $POMDP^*_{\alpha\beta}$ state transition function $T^*$ and observation function $O^*$ presented in Table 10 and Table 11, respectively. Those functions specify the $POMDP^*_{\alpha\beta}$ for regulating this "new" social exchange strategy, for a specific $*$.

## 8 Simulation of strategy-based social exchanges

To analyze the mechanism for recognizing and learning social exchange strategies in interactions that happen in a simulated open society, we consider the self-regulating agent $\alpha$ interacting with a multi-strategy agent $\beta$, so that $\alpha$ could experiment the need to regulate different exchange strategies.

These two agents are as follows:

*The multi-strategy (MS) agent* $\beta$: The MS agent $\beta$ is able to adopt different social exchange strategies. For simplicity, the MS agent chooses randomly the strategy to be adopted in each exchange process, which is composed by a fixed set of exchange stages, and uses the same strategy until the processes ends.

*The self-regulating (SR) agent* $\alpha$: The SR agent is a BDI-POMDP-HMM agent that has to recognize or learn the

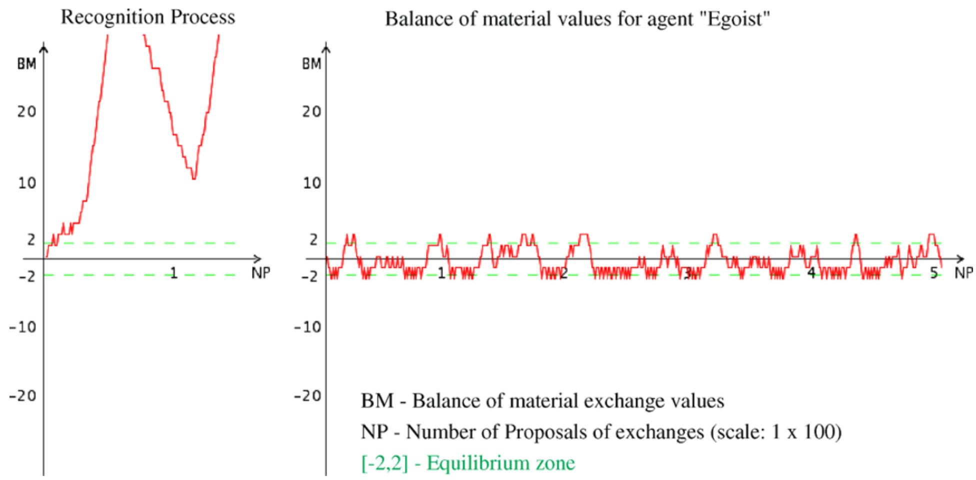



**Fig. 1** Recognizing and regulating an Egoistic strategy-based agent

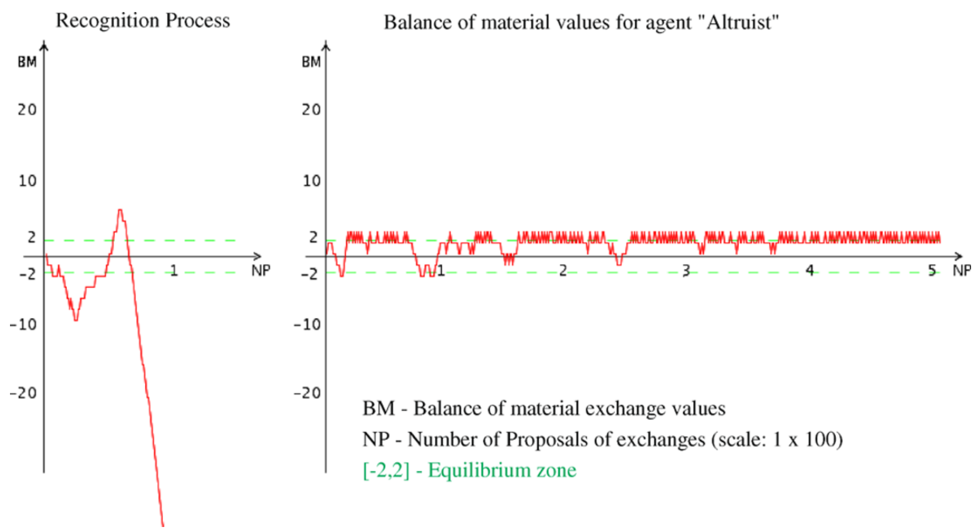



**Fig. 2** Recognizing and regulating an Altruistic strategy-based agent

social exchange strategy adopted by $\beta$, as explained in Sects. 6 and 7, and then proceed to the regulation of the exchanges, as explained in Sect. 5.

Any exchange process always starts in the equilibrium state (material balance in the interval $[-2; 2]$).

In the simulations, we consider exchange processes composed by a set of 650 proposals of exchange stages. The first 150 proposals are used by the SR agent $\alpha$ for the process of recognition/learning the social exchange strategy adopted by the MS agent $\beta$, observing its behavior in each state (balance of material exchange values). For that, the SR agent $\alpha$ performs four sequences of exchanges proposals:

1. 30 proposals of exchange stages of type $\mathrm{I}_{\alpha\beta}$ or $\mathrm{II}_{\alpha\beta}$, alternatively, i.e., alternating the proposals of `do-service` and `ask-service` actions, in order to analyze $\beta$'s exchange behavior around the equilibrium state.
2. 30 proposals of exchange stages of type $\mathrm{I}_{\alpha\beta}$, i.e., only proposing `do-service` actions, in order to observe $\beta$'s behavior in favorable states;
3. 60 proposals of exchange stages of type $\mathrm{II}_{\alpha\beta}$, i.e., only proposing `ask-service` actions, leading the agent $\beta$'s to unfavorable states, passing through the equilibrium state, in order to observe $\beta$'s behavior in unfavorable states;
4. 30 proposals of exchange stages of type $\mathrm{I}_{\alpha\beta}$ in order to lead the agent $\beta$ back again to the starting state (the equilibrium state).

The implementation of the strategy-based social exchange simulator [38] was done in *AgentSpeak* using *Jason* [5].

### 8.1 Recognizing and regulating known social exchange strategies

The graphics on the left of Figs. 1, 2 and 3 show examples of processes of recognition of already known social exchange strategies, the egoistic, altruistic and tolerant strategies, respectively.

**Fig. 3** Recognizing and regulating a Tolerant strategy

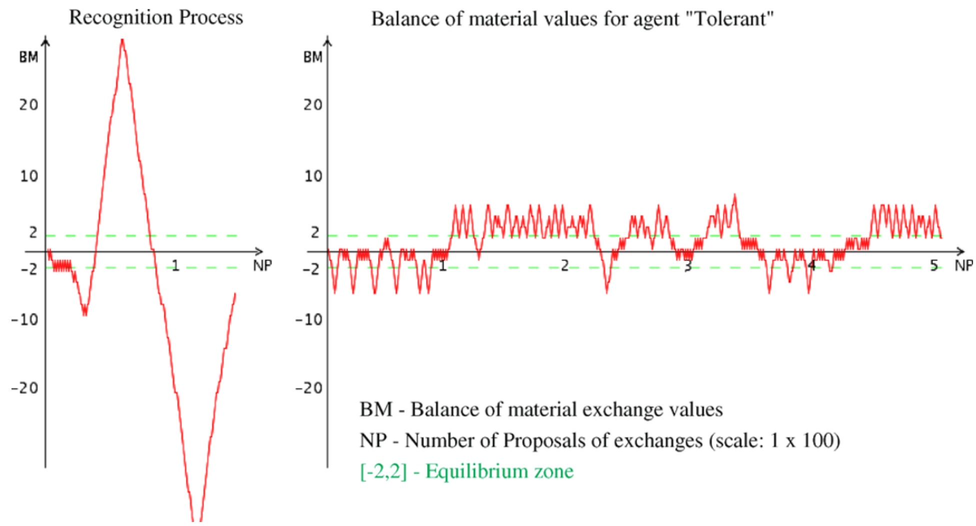
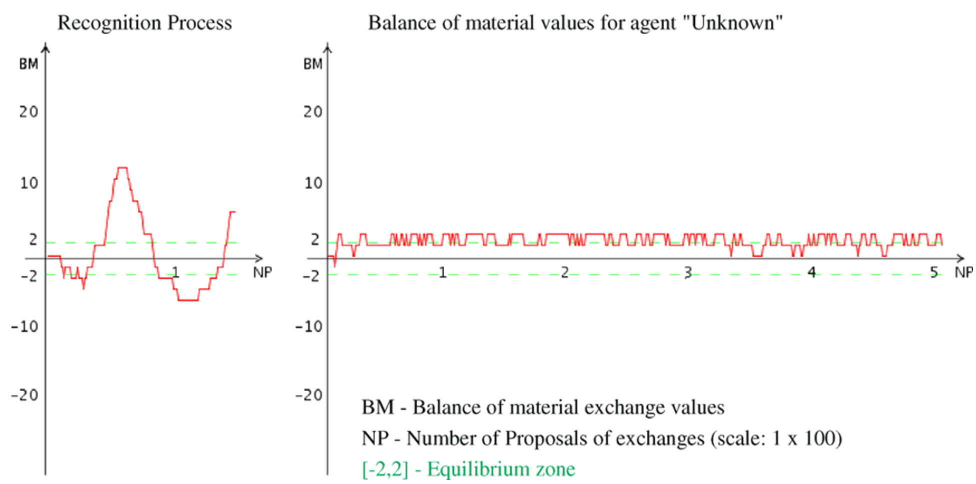



**Fig. 4** Learning and regulating an unknown strategy




Analyzing Fig. 1 (left) and Fig. 2 (left), e.g., it is possible to observe the opposite exchange behaviors presented by egoistic and altruistic strategy-based agents $\beta$ during the recognition processes conducted by $\alpha$. The tendency of an egoistic strategy-based agent $\beta$ (Fig. 1 (left)) was to look for favorable material balances (positive zone), meaning that it prefers exchange stages of type $I_{\alpha\beta}$, when the agent $\alpha$ performs a service to it. On the contrary, altruistic strategy-based agents $\beta$ (Fig. 2 (left)) preferred to stay in unfavorable states (negative zone), that is, it opted for exchange stages of type $II_{\alpha\beta}$, when it performs a service to the agent $\alpha$.

Observe in Fig. 3 (left) that, in the recognition process of the tolerant exchange strategy, the agent $\beta$ has visited all states (from unfavorable to favorable states, passing through the equilibrium zone), since its IERs are very low, and then, in general, it has accepted most of the exchange proposals.

The graphics on the right of Figs. 1, 2 and 3 show that the SR agent was able to recognize the different strategies and to deliberate on the adequate plans for regulating the interactions around the equilibrium zone.

## 8.2 Learning and regulating new social exchange strategies

The graphics on the left of Figs. 4, 5 and 6 show examples of processes of learning unknown social exchange strategies. In those processes, the SR agent was not able to recognize the strategies adopted by the MS agent, since their IERs were not contained in any range of IERs in its knowledge base.

Then the SR agent used its learning mechanism based in HMMs to build the POMDP model in order to extract the BDI plans for each new social exchange strategy, registering new ranges of IERs for future recognizing processes.

The graphics on the left of Figs. 4, 5 and 6 show that the agent succeeded in constructing the adequate plans for regulating the interactions.

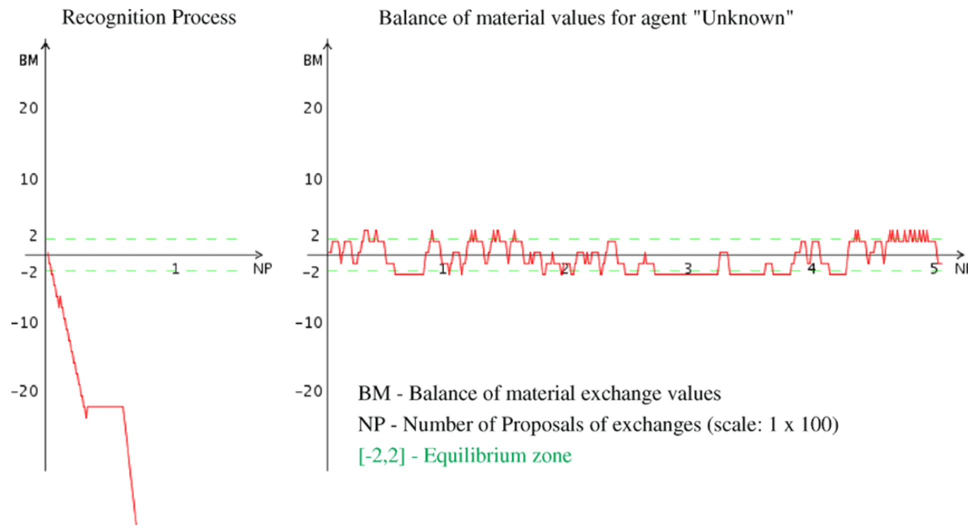



**Fig. 5** Learning and regulating unknown strategies

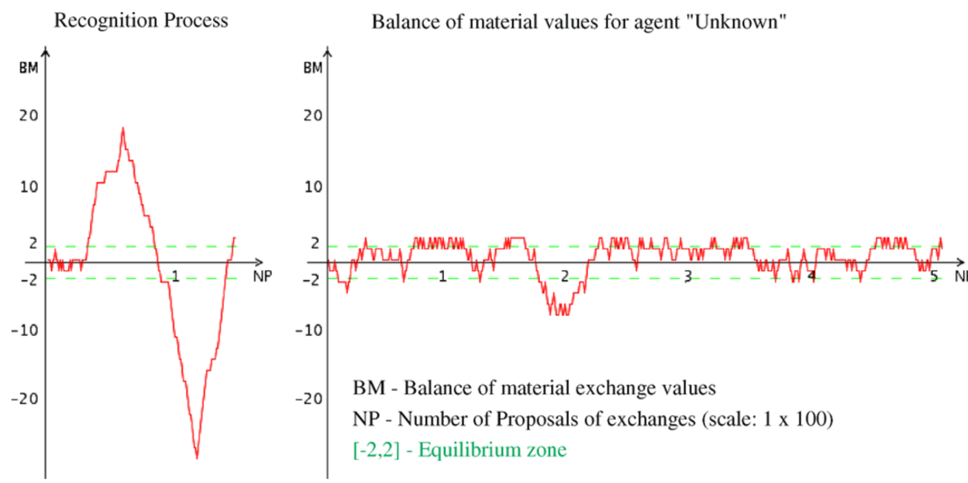



**Fig. 6** Learning and regulating unknown strategies

## 9 Related work on social exchanges in multiagent systems

As discussed in the Introduction, our modeling of agent interactions based on the Piaget's Theory of Social Exchanges was first proposed in 2005 [16], and, since then, our many works on this subject have pointed out the problems involved in the self-regulation of strategy-based social exchanges in open multiagent systems [12–15, 38–40]. To the best of our knowledge, there is no other work by different authors concerned with this subject.

However, social exchanges have been also considered in different other agent scenarios, although the problem of the equilibrium of the social exchanges has been always present, since this is one of the main concerns of Piaget's theory.

One such line of works is due to M. Rodrigues in cooperation with different partners [45–47, 49]. In Rodrigues and Costa [45] and Rodrigues et al. [46], an initial work on an algebra of exchange values was introduced for the modeling of social interactions in agent societies, together with a social-reasoning mechanism and the specification of structures for storing and manipulating such values, presenting an example of a political process of lobbying through campaign contributions. However, the approach to the exchange values proposed in those works was not qualitative, as neither was the social-reasoning mechanism. Then, in [16], we introduced an algebra of qualitative exchange values, resulting in our initial approach for manipulating exchange values and reasoning about equilibrated social exchanges, reinterpreting the politician/voters scenario [12, 15] first analyzed by Rodrigues et al. [45, 46].

Rodrigues and Luck [44, 47–50] introduced a rich approach based on the Theory of Social Exchanges for the modeling of interactions in open multiagent systems, presenting a system for analysing/evaluating partner selection and cooperative interactions in the Bioinformatics domain, which is characterized by frequent, extensive and dynamic exchanges of services.

The works by Grimaldo et al. [23–27] presented an interesting application of the Theory of Social Exchanges in the coordination of intelligent virtual agents and sociability in a virtual university bar scenario (in a 3D dynamic environment), modeled as a market-based social model, where groups of different types of waiters (e.g., coordinated, social, egalitarian) and customers (e.g., social, lazy) interact with both the objects in the scene and the other virtual agents. In [26], they presented a multi-modal agent decision making model, called MADeM, in order to provide virtual agents with socially acceptable decisions, coordinated social behaviors (e.g., task passing or planned meetings), based on the evaluation of the social exchanges.

In Franco et al. [20, 21], social exchange values are used to support arguments about the assessment of exchanges. Together with the power-to-influence social relationship, those arguments were also used to help the agents to decide about the continuation or the interruption of on-going interactions.

Social exchanges were also considered to model interactions in the Population-Organization model (PopOrg) [10, 11], a formal model for studying the organization of open multiagent systems and their functional and structural dynamics. In the work by Barbosa and Costa [2], on the other hand, social exchanges were model as processes of the CSP [28] formal language for concurrent processes.

## 10 Conclusion

This paper introduced a hybrid BDI-POMDP-HMM agent model for the self-regulation of social exchanges in open societies.

We extended the model presented in previous work [38], in order to solve the problems of the recognition and/or learning of social exchange strategies.

For the recognition problem, the self-regulating agent observes the social exchange behavior of the partner agent in order to compute its IERs, which are indices that represent patterns of refusals to exchange proposals, observed in the agents' strategies.

For the learning problem, the self-regulating agent uses HMM algorithms for learning the probabilistic state transition and observation functions of the global observable behavior of any unknown social exchange strategy. Those functions are then translated into the action-based state transition and observation functions, in order to define a new social exchange strategy POMDP model, from which optimal policy the algorithm **policyToBDIplans** extracts BDI plans.

The translation algorithms were defined after formally integrating the HMM and the POMDP models. This integration was possible since there is an isomorphism between the sets of states of the models, although this isomorphism could be established only because the HMM states are associated to stimulus from the environment (in this case, exchange proposals from the self-regulating agent).

The various simulations that we produced showed that the proposed approach is viable and may be a good solution in contextualized applications based on the theory of social exchanges, such as the ones discussed in Sect. 9.

Future work will be concerned with the simulation of organizations such as research institutes, hotels or firms, which are organizational environments rich in non-economic service exchanges (e.g., between department colleagues, between manager and secretaries, between costumers and employees), in order to help the analysis of the consequences of equilibrated/non-equilibrated social interaction and workers' reciprocity for optimal organizational design [18, 32].

**Acknowledgements** The work is supported by the following Brazilian funding agencies: CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), under the Proc. No. 305131/10-9, 483257/09-5, 560118/10-4, 304580/07-4, and CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior). We are very grateful to the anonymous referees for their valuable comments and suggestions that helped us to improve the paper.

## References

1. Aaron E, Admoni H (2010) Action selection and task sequence learning for hybrid dynamical cognitive agents. Robot Auton Syst 58(9):1049–1056
2. Barbosa RM, Costa ACR (2010) Using CSP in the formal specification of the micro-organizational level of multiagent systems. In: Trappl R (ed) Cybernetics and systems 2010, Proceedings of the 20th European meeting on cybernetics and systems research, EMCRC 2010. Austrian Society for Cybernetic Studies, Vienna, pp 459–464
3. Bernstein DS, Givan R, Immerman N, Zilberstein S (2002) The complexity of decentralized control of MDPs. Math Oper Res 27(4):819–840
4. Blau P (1964) Exchange and power in social life. Wiley, New York
5. Bordini RH, Hübner JF, Wooldrige M (2007) Programming multi-agent systems in AgentSpeak using Jason. Wiley series in agent technology. Wiley, Chichester
6. Carbonell JG (1980) Towards a process model of human personality traits. Artif Intell 15(1, 2):49–74
7. Castelfranchi C, Rosis F, Falcone R, Pizzutilo S (1997) A testbed for investigating personality-based multiagent cooperation. In: Proceedings of the symposium on logical approaches to agent modeling and design. Aix-en-Provence
8. Castelfranchi C, Rosis F, Falcone R, Pizzutilo S (1998) Personality traits and social attitudes in multiagent cooperation. Appl Artif Intell 12:649–675
9. Cook K, Yamaguchi T (1990) Power relations in exchange networks. Am Sociol Rev 55:297–300
10. Costa ACR, Dimuro GP (2009) Introducing social groups and group exchanges in the PopOrg model. In: Proceedings of AAMAS 2009, vol 1. IFAAMAS, Budapest, pp 1297–1298
11. Costa ACR, Dimuro GP (2009) A minimal dynamical organization model. In: Dignum V (ed) Handbook of research on multi-agent systems: semantics and dynamics of organizational models. IGI Global, Hershey, pp 419–445

12. Dimuro GP, Costa ACR (2006) Exchange values and self-regulation of exchanges in multi-agent systems: the provisory, centralized model. In: Brueckner S, Serugendo GDM, Hales D, Zambonelli F (eds) Engineering self-organising systems. LNCS, vol 3910. Springer, Berlin, pp 75–89
13. Dimuro GP, Costa ACR (2006) Interval-based Markov decision processes for regulating interactions between two agents in multi-agent systems. In: Dongarra J, Madsen K, Wasniewski J (eds) Applied parallel computing. LNCS, vol 3732. Springer, Berlin, pp 102–111
14. Dimuro GP, Costa ACR, Gonçalves LV, Hübner A (2006) Regulating social exchanges between personality-based non-transparent agents. In: Gelbukh AF, Garça CAR (eds) MICAI 2006: Advances in artificial intelligence. LNCS, vol 4293. Springer, Berlin, pp 1105–1115
15. Dimuro GP, Costa ACR, Gonçalves LV, Hübner A (2007) Centralized regulation of social exchanges between personality-based agents. In: Coordination, organizations, institutions, and norms in agent systems II. LNCS, vol 4386. Springer, Berlin, pp 338–355
16. Dimuro GP, Costa ACR, Palazzo LAM (2005) Systems of exchange values as tools for multi-agent organizations. J Braz Comput Soc 11(1):31–50
17. Dimuro GP, Santos AV, Bedregal BC, Costa ACR (2009) Fuzzy evaluation of social exchanges between personality-based agents. In: Lopes LS, Lau N, Mariano P, Rocha LM (eds) New trends in artificial intelligence, Proceedings of the 14th Portuguese conference on artificial intelligence. APIA, Aveiro, pp 451–462
18. Dur R, Roelfsema H (2010) Social exchange and common agency in organizations. J Socio-Econ 39(1):55–63
19. Emerson R (1976) Social exchange theory. Annu Rev Sociol 2:335–362
20. Franco MHI, Costa ACR, Coelho H (2010) Exchange values and social power supporting the choice of partners. Pueblos Fronteras Digital 6(9) (Available at http://www.pueblosyfronteras.unam.mx/, accessed in February, 2010)
21. Franco MHI, Costa ACR, Coelho H (2011) Simulating argumentation about exchange values in multi-agent interactions. In: Dimuro GP, Costa ACR, Sichman JS, Tedesco P, Adamatti DF, Balsa J, Antunes L (eds) Advances in social simulation, Post-proceedings of the Brazilian workshop on social simulation. IEEE, Los Alamitos, pp 105–112
22. Ghasem-Aghaee N, Ören TI (2003) Towards fuzzy agents with dynamic personality for human behavior simulation. In: Proceedings of the 2003 summer computer simulation conference, Montreal, July 20–24, 2003. SCS, San Diego, pp 3–10
23. Grimaldo F, Lozano M, Barber F (2007) Coordination and sociability for intelligent virtual agents. In: Sichman J, Noriega P, Padget J, Ossowski S (eds) Coordination, organizations, institutions, and norms in agent systems III. LNCS, vol 4870. Springer, Berlin, pp 58–70
24. Grimaldo F, Lozano M, Barber F (2007) A multiagent framework to animate socially intelligent agents. In: Innovations in hybrid intelligent systems. Advances in intelligent and soft computing, vol 44. Springer, Berlin, pp 17–24
25. Grimaldo F, Lozano M, Barber F (2007) Social animation in complex environments. In: Intelligent virtual agents. LNCS, vol 4722. Springer, Berlin, pp 379–380
26. Grimaldo F, Lozano M, Barber F (2008) Madem: a multi-modal decision making for social MAS. In: Proceedings of the 7th international joint conference on autonomous agents and multiagent systems, AAMAS'08, Estoril, May 2008. IFAAMAS, Richland, pp 183–190
27. Grimaldo F, Lozano M, Barber F, Vigueras G (2008) Simulating socially intelligent agents in semantic virtual environments. Knowl Eng Rev 23(4):369–388
28. Hoare CAR (1985) Communicating sequential processes. Prentice-Hall, Englewood Cliffs
29. Homans G (1961) Social behavior—its elementary forms. Harcourt, Brace & World, New York
30. Kaelbling LP, Littman ML, Cassandra AR (1998) Planning and acting in partially observable stochastic domains. Artif Intell 101(1–2):99–134
31. Kamphorst B, van Wissen A, Dignum V (2009) Incorporating BDI agents into human-agent decision making research. In: Aldewereld H, Dignum V, Picard G (eds) Engineering societies in the agents world X. LNCS, vol 5881. Springer, Berlin, pp 84–97
32. Ma E, Qu H (2011) Social exchanges as motivators of hotel employees' organizational citizenship behavior: the proposition and application of a new three-dimensional framework. Int J Hosp Manag 30(3):680–688
33. Nair R, Tambe M (2005) Coordinating teams in uncertain environments: a hybrid BDI-POMDP approach. In: Bordini R, Dastani M, Dix J, Seghrouchni A (eds) Programming multi-agent systems. LNCS, vol 3346. Springer, Berlin, pp 1–37
34. Nair R, Tambe M (2005) Hybrid BDI-POMDP framework for multiagent teaming. J Artif Intell Res 23:367–420
35. Oliehoek F, Visser A (2010) A decision-theoretic approach to collaboration: principal description methods and efficient heuristic approximations. In: Babuška R, Groen F (eds) Interactive collaborative information systems. Studies in computational intelligence, vol 281. Springer, Berlin, pp 87–124
36. Ören TI, Ghasem-Aghaee N (2003) Personality representation processable in fuzzy logic for human behavior simulation. In: Proceedings of the 2003 summer computer simulation conference, Montreal, July 20–24, 2003. SCS, San Diego, pp 11–18
37. Paruchuri P, Bowring E, Nair R, Pearce J, Schurr N, Tambe M, Varakantham P (2006) Multiagent teamwork: hybrid approaches. In: Computer society of India communications. CSI (Invited Talk, available at http://teamcore.usc.edu/publications.htm)
38. Pereira D, Gonçalves LV, Dimuro GP, Costa ACR (2008) Towards the self-regulation of personality-based social exchange processes in multiagent systems. In: Zaverucha G, Costa AL (eds) Advances in artificial intelligence. LNAI, vol 5249. Springer, Berlin, pp 113–123
39. Pereira DR, Gonçalves LV, Dimuro GP, Costa ACR (2008) Constructing BDI plans from optimal POMDP policies, with an application to agentspeak programming. In: Proceedings of conferencia Latinoamerica de informática, CLEI'08. SADIO, Santa Fe, pp 240–249
40. Pereira DR, Gonçalves LV, Dimuro GP, Costa ACR (2011) An analysis of two regulation mechanisms for personality-based social exchange processes. In: Dimuro GP, Costa ACR, Sichman JS, Tedesco P, Adamatti DF, Balsa J, Antunes L (eds) Advances in social simulation, post-proceedings of the Brazilian workshop on social simulation. IEEE, Los Alamitos
41. Piaget J (1995) Sociological studies. Routlege, London
42. Rabiner LR (1989) A tutorial on hidden Markov models and selected applications in speech recognition. Proc IEEE 77(2):257–286
43. Rao AS, Georgeff MP (1991) Modeling rational agents within a BDI-architecture. In: Fikes R, Sandewall E (eds) Proceedings of the 2nd international conference on principles of knowledge representation and reasoning. Morgan Kaufmann, San Mateo, pp 473–484
44. Rodrigues MR (2007) Social techniques for effective interactions in open cooperative systems. PhD thesis, University of Southampton, Southampton
45. Rodrigues MR, Costa ACR (2004) Using qualitative exchange values to improve the modelling of social interactions. In: Hales D, Edmonds B, Norling E, Rouchier J (eds) Proceedings of IV workshop on agent based simulations, MABS'03, Melbourne, 2003. LNAI, vol 2927. Springer, Berlin, pp 57–72

46. Rodrigues MR, Costa ACR, Bordini R (2003) A system of exchange values to support social interactions in artificial societies. In: Proceedings of the 2nd international conference on autonomous agents and multiagents systems, AAMAS'03. ACM Press, Melbourne, pp 81–88
47. Rodrigues MR, Luck M (2006) Analysing partner selection through exchange values. In: Antunes L, Sichman J (eds) Proceedings of the VI workshop on agent based simulations, MABS'05, Utrecht, 2005. LNAI, vol 3891. Springer, Berlin, pp 24–40
48. Rodrigues MR, Luck M (2006) Evaluating dynamic services in bioinformatics. In: Proceedings of the tenth international workshop on cooperative information agents, 2006, Edinburgo. LNAI, vol 4149. Springer, Berlin, pp 183–197
49. Rodrigues MR, Luck M (2007) Cooperative interactions: an exchange values model. In: Noriega P, Vázquez-Salceda J, Boella G, Boissier O, Dignum V, Formara N, Matson E (eds) Coordination, organizations, institutions and norms in agent systems II. LNAI, vol 4386. Springer, Berlin, pp 356–371
50. Rodrigues MR, Luck M (2009) Effective multiagent interactions for open cooperative systems rich in services. In: Sierra C, Castelfranchi C, Decker KS, Sichman JS (eds) Proceedings of the 8th international joint conference on autonomous agents and multiagent systems, AAMAS'09, Budapest, May 2009. IFAAMAS, Richland, pp 1273–1274
51. Schut M, Wooldridge M, Parsons S (2001) Reasoning about intentions in uncertain domains. In: Benferhat S, Besnard P (eds) Symbolic and quantitative approaches to reasoning with uncertainty. LNCS, vol 2143. Springer, Berlin, pp 84–95
52. Schut MC, Wooldridge M, Parsons S (2002) On partially observable MDPs and BDI models. In: d'Inverno M, Luck M, Fisher M, Preist C (eds) Foundations and applications of multi-agent systems, Selected papers from the UKMAS workshop. LNCS, vol 2403. Springer, Berlin, pp 243–260
53. Simari GI, Parsons S (2006) On the relationship between MDPs and the BDI architecture. In: Nakashima H, Wellman MP, Weiss G, Stone P (eds) Proceedings of the 5th international joint conference on autonomous agents and multiagent systems, Hakodate. ACM, New York, pp 1041–1048
54. Trigo P, Coelho H (2010) Decision making with hybrid models: the case of collective and individual motivations. Int J Reason-based Intell Syst 2(1):60–72
55. Trigo P, Coelho H (2011) Simulating a multi-agent electricity market. In: Dimuro GP, Costa ACR, Sichman JS, Tedesco P, Adamatti DF, Balsa J, Antunes L (eds) Advances in social simulation, post-proceedings of the Brazilian workshop on social simulation. IEEE, Los Alamitos, pp 91–96
56. Wooldridge M (2000) Reasoning about rational agents. Intelligent robots and autonomous agents. MIT Press, Cambridge
57. Zafirovski M (2003) Some amendments to social exchange theory: a sociological perspective. Theory Sci 4(2) (Available at http://theoryandscience.icaap.org/, accessed in February, 2011)