Centralized Regulation of Social Exchanges Between Personality-Based Agents

Graçaliz Pereira Dimuro, Antônio Carlos da Rocha Costa, Luciano Vargas Gonçalves, and Alexandre Hübner

Escola de Informática, PPGINF, Universidade Católica de Pelotas
96010-000 Pelotas, Brazil
{liz,rocha,llvarga,hubner}@ucpel.tche.br

Abstract. This paper presents a centralized mechanism for solving the coordination problem of personality-based multiagent systems from the point of view of social exchanges. The agents may have different personality traits, which induce different attitudes towards both the regulation mechanism and the possible profits of social exchanges. A notion of exchange stability can be defined, and the connections between agents' personalities and deviations of social exchanges from the stability point can be established. The model supports a decision procedure based on Qualitative Interval Markov Decision Processes, that can solve the problem of keeping the stability of social exchanges, in spite of the different personality traits of the agents. The paper deals only with transparent agents (agents that allow the external access to their balances of exchange values), but we hint on the case of non-transparent agents. The model is analyzed theoretically and contextualized simulations are presented.

1 Introduction

Social control is a powerful notion for explaining the self-regulation of a society, and the various possibilities for its implementation have been considered, both in natural and artificial societies [1,2]. As mentioned in [3], social control, or coordination mechanisms, vary according to the structure of the society: hierarchy, market or network-oriented societies tend to coordinate activities through, respectively, authority supervision, price mechanism or collaboration mechanisms. Our work aims at the simulation of network-oriented societies with collaboration based social control. However, up to now, we are dealing with a hierarchical model, and the system of exchange values that constitutes the basis of the social control model that we adopted seems to be a price mechanism, although one based on qualitative values, as we show presently.

The centralized social control mechanism that we introduced in [4], concerning small social groups, is based on the Piaget's theory of *exchange values* [5], where a variety of social norms (moral, juridical, even economic rules) are rooted in the qualitative economy of exchange values that emerges when individuals evaluate their interactions. That control mechanism is performed by a *social equilibrium supervisor* that, at each time, decides on which actions it should recommend

P. Noriega et al. (Eds.): COIN 2006 Workshops, LNAI 4386, pp. 338–355, 2007.

agents to perform in order to lead the system towards the equilibrium, regarding the balance of the exchange values involved in their exchanges.

The qualitative exchange values are represented using techniques of Interval Mathematics [6]. The *equilibrium point* of the exchanges between a pair of agents is defined as a pair of intervals, each enclosing the value zero, meaning that benefits and losses in exchanges, for each agent, compensate one another. Then, the equilibrium supervisor builds on *Qualitative Interval Markov Decision Processes* (QI-MDP), where states are represented by intervals, actions are interval operations [6], and equality of intervals is interpreted as a loose equivalence relation (two intervals are equivalent if their midpoints are "approximately" equal).

In general, however, since the agents may have different objectives, it may happen that the exchange balance of a given agent, regarding its exchanges with another agent, becomes stable (after a certain period of time) around a value different from zero. That is, in general, agents stabilize their exchanges in non-equilibrated ways, thus keeping the society disequilibrated, as a whole. Given two agents, the *pair of exchange values* in which they stabilize their respective exchange balances is called the *stability point* of the exchanges between them. Such stability point may vary with time.

In this paper, trying to advance the development of a future model of *decentralized* social control, we extend the centralized control architecture presented in [7], in order to consider a society with *personality-based* agents. We propose a social control mechanism coordinated by a *stability supervisor*, whose duty is: (i) to determine, at each time, the target stability point for each pair of agents in the system (which is not necessarily around the value zero); (ii) to decide which actions should be recommended for each pair of agents in order to lead them towards that stability point; (iii) to maintain them stable around that point, until (for some reason) another stability point for some pair of agents is required.

As explained in [8], a realistic account of agent interactions has to consider that agents may have different *interaction personalities*, in order to allow for the agents to participate in different ways in social interactions, depending not only on the way tasks were delegated to them, but also on the way the agents assess their own contributions and the contributions of the other agents to the interaction. So, in this paper, we allow for the agents to have different personality traits, which induce different attitudes towards the social control mechanism (blind obedience, eventual obedience etc.) and the possible profits of social exchanges (egoism, altruism etc.). As a consequence, the agents may or may not follow the recommendations given by the stability supervisor, thus creating a probabilistic social environment, from the point of view of the social control.

Also, we allow the agents to control the supervisor access to their internal states, behaving either as *transparent agents* (agents that allow full external access to their internal states) or as *non-transparent agents* (agents that restrict such external access). In the paper, however, we focus on the supervisor dealing only with transparent agents. Then, it has full knowledge of the agents' personality traits and has access to all current balances of exchange values, and so it is able to choose, at each step, the adequate recommendation for each agent.

We note, however, that the motivation for establishing a social control mechanism (for instance, social stability, social equilibrium or disequilibrium etc.), is usually not inscribed in the details of the social control mechanism itself. That motivation usually lies in the agents themselves or in the application context of the system. Thus, the social control mechanisms that we are developing are neutral with respect to those motivations, serving any of those purposes.

Section 2 shows our modeling of social exchanges. The regulation mechanism of exchanges is introduced in Sect. 3. Section 4 presents the QI-MDP model for the regulation of exchanges between transparent personality-based agents, and the stability analysis. Section 5 shows a sample simulation. Related work is discussed in Sect. 6. Section 7 is the Conclusion and discussion of further work.

2 The Modelling of Social Exchanges

According to Piaget's approach [5], the evaluation of an exchange by an agent is done on the basis of a *scale* of *exchange values* (that are of a qualitative, subjective nature, like those everyone uses to judge the daily exchanges he has: *good*, *bad*, *better than* etc.). In general, those values cannot be faithfully represented quantitatively, due to the lack of neat objective conditions for their measurement. Then, following the approach introduced in [4,9], techniques from Interval Mathematics [6] are used to represent any exchange value[1] as a real interval $X = [x_1, x_2] = \{x \in \mathbb{R} \mid x_1 \leq x \leq x_2\}$, with $-L \leq x_1 \leq x \leq x_2 \leq L$, x_1, x_2, for a bound $L \in \mathbb{R}$, $L > 0$. The set of such intervals is denoted by $\mathbb{IR}_L$.

Analogously to [4,7], consider a reference value h (an anchor for the stability point) such that $-L < h < L$, and a tolerance $\epsilon \in \mathbb{R}_+$. We build an *h-centered scale of exchange values* as an algebraic structure $\langle \mathbb{IR}_L, +, \mathbf{X}_h, ' \rangle$, where: (i) the *L-bounded interval addition* operation $+$ is well defined; (ii) $\mathbf{X}_h = \{X \in \mathbb{IR}_L \mid mid(X) \in [h-\epsilon, h+\epsilon]\}$ is the set of *h-reference intervals*, where $mid(X) = \frac{x_1+x_2}{2}$ is the mid point of X; (iii) an *h-compensation interval* of an interval $X \in \mathbb{IR}_L$ is any interval $X' \in \mathbb{IR}_L$ such that $X + X'$ is an h-reference interval; (iv) the least compensation interval of X is given by $[-mid(X) + h - \epsilon, -mid(X) + h + \epsilon]$.

A *social exchange* between two agents, α and β, involves two types of stages. In stages of type $\mathrm{I}_{\alpha\beta}$, α realizes an action on behalf of (a "service" for) β. The *exchange values* involved in this stage are the following: $r_{\mathrm{I}_{\alpha\beta}}$, which is the value of the *investment* done by α for the realization of a service for β (this value is always *negative*); $s_{\mathrm{I}_{\beta\alpha}}$, which is the value of β's *satisfaction* due to the receiving of the service done by α; $t_{\mathrm{I}_{\beta\alpha}}$ is the value of β's *debt*, the debt it acquired to α for its satisfaction with the service done by α; and $v_{\mathrm{I}_{\alpha\beta}}$, which is the value of the *credit* that α acquires from β for having realized the service for β. In stages of type $\mathrm{II}_{\alpha\beta}$, α asks the payment for the service previously done for β, and the values related with this exchange have similar meaning.

The values $r_{\mathrm{I}_{\alpha\beta}}$, $s_{\mathrm{I}_{\beta\alpha}}$, $r_{\mathrm{II}_{\beta\alpha}}$ and $s_{\mathrm{II}_{\alpha\beta}}$ are called *material values* (investments and satisfactions), generated by the evaluation of *immediate exchanges*; the

[1] Our choice makes the representation operational and the decision process computationally viable, without being unfaithful to Piaget's approach [4].

values $t_{\mathrm{I}_{\beta\alpha}}$, $v_{\mathrm{I}_{\alpha\beta}}$, $t_{\mathrm{II}_{\beta\alpha}}$ and $v_{\mathrm{II}_{\alpha\beta}}$ are the *virtual values* (credits and debts), concerning *deferred exchanges*, which are expected to happen in the future [4,5].

The exchange values are undefined if either no service is done in a stage I, or no credit is charged in a stage II. Also, it is not possible for α to realize a service for β and, at the same, to charge him a credit. A *configuration* of ex-values is specified by one of the tuples of well defined values: $(r_{\mathrm{I}_{\alpha\beta}}, s_{\mathrm{I}_{\beta\alpha}}, t_{\mathrm{I}_{\beta\alpha}}, v_{\mathrm{I}_{\alpha\beta}})$, $(r_{\mathrm{I}_{\beta\alpha}}, s_{\mathrm{I}_{\alpha\beta}}, t_{\mathrm{I}_{\alpha\beta}}, v_{\mathrm{I}_{\beta\alpha}})$, $(v_{\mathrm{II}_{\alpha\beta}}, t_{\mathrm{II}_{\beta\alpha}}, r_{\mathrm{II}_{\beta\alpha}}, s_{\mathrm{II}_{\alpha\beta}})$, $(v_{\mathrm{II}_{\beta\alpha}}, t_{\mathrm{II}_{\alpha\beta}}, r_{\mathrm{II}_{\alpha\beta}}, s_{\mathrm{II}_{\beta\alpha}})$.

A *social exchange process* is composed by a sequence of stages of type $\mathrm{I}_{\alpha\beta}$ and/or $\mathrm{II}_{\alpha\beta}$ in a set of discrete instants of time. The *material results*, according to the points of view of α and β, are given by the sum of the well defined material values involved in the process, and are denoted, respectively, by $\mathbf{m}_{\alpha\beta}$ and $\mathbf{m}_{\beta\alpha}$. The *virtual results* $\mathbf{v}_{\alpha\beta}$ and $\mathbf{v}_{\beta\alpha}$ are defined analogously.

A *stability point* is a pair of balances of exchanges that is desired that a pair of agents should maintain for a certain period of time, established according to conditions and constraints imposed by the system's external and internal environments (see Sect. 3). A social exchange process is said to be in *material stability* if in all its duration it holds that the pair of material results $(\mathbf{m}_{\alpha\beta}, \mathbf{m}_{\beta\alpha})$ encloses a given stability point $(\omega_{\alpha\beta}, \omega_{\beta\alpha}) \in \mathbb{R} \times \mathbb{R}$. It is said in *material equilibrium*[2] if both $\mathbf{m}_{\alpha\beta}$ and $\mathbf{m}_{\beta\alpha}$ enclose the zero.

Let $H = \{-L, -L + \frac{L}{n}, -L + 2\frac{L}{n}, \ldots, L - 2\frac{L}{n}, L - \frac{L}{n}, L\}$ be the set of possible reference values induced on $\mathbb{IR}_L$ by a given $n \in \mathbb{N}^*_+$, and $\kappa_n = \frac{L}{n}$ be the *accuracy* of the stability supervisor. Given a target stability point $(\omega_{\alpha\beta}, \omega_{\beta\alpha}) \in \mathbb{R} \times \mathbb{R}$ for the exchange process between the pair of agents α and β, occurring during a certain period of time, a pair of reference values $(h_{\alpha\beta}, h_{\beta\alpha}) \in H \times H$ is chosen such that $\omega_{\alpha\beta} \in [h_{\alpha\beta} - \epsilon, h_{\alpha\beta} + \epsilon]$ and $\omega_{\beta\alpha} \in [h_{\beta\alpha} - \epsilon, h_{\beta\alpha} + \epsilon]$, for a tolerance $0 < \epsilon < \frac{L}{n}$ and machine numbers $h_{\alpha\beta} \pm \epsilon$, $h_{\beta\alpha} \pm \epsilon$. The stability supervisor builds two scales of exchange values, one that is $h_{\alpha\beta}$-centered (for the agent α) and other that is $h_{\beta\alpha}$-centered (for the agent β). The index $\alpha\beta$ ($\beta\alpha$) of a reference value $h_{\alpha\beta}$ ($h_{\beta\alpha}$) will be omitted when it is not relevant in the context.

3 The Social Exchange Regulation Mechanism

Figure 1 shows the architecture of our social exchange regulation mechanism, which extends the one proposed in [7] with (i) a module for the evaluation of stability points and (ii) a learning module based on Hidden Markov Models (HMM) [10]. The *stability supervisor*, at each time, uses an *Evaluation Module* to analyze the conditions and constraints imposed by the system's external and internal environments (not shown in the figure), determining the target equilibrium point. To regulate *transparent* agents, the supervisor uses two *Balance Modules*, $\Sigma_{material}$ and $\Sigma_{virtual}$, to calculate their material and virtual results of the performed exchanges. To regulate *non-transparent* agents, the supervisor uses a *HMM Module* to observe their behavior in exchanges and then

[2] 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.

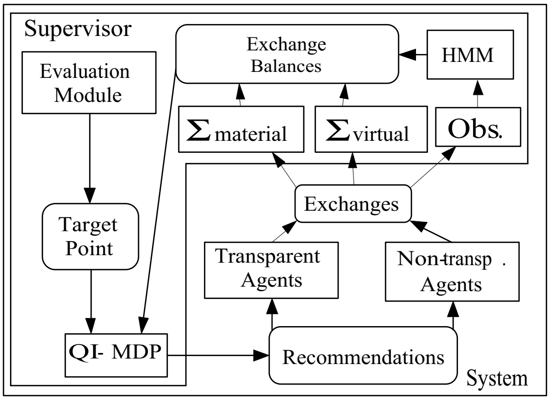

Fig. 1. The social exchange regulation mechanism

recognize and maintain an adequate model of the personality traits of such agents, generating *plausible balances* of their material exchange values.

Taking both the directly observed and the indirectly calculated material results, together with the currently target stability point, the supervisor uses the module that implements a personality-based QI–MDP to decide on recommendations of exchanges for the two agents[3], in order to keep the material results of exchanges stable. It also takes into account the virtual results of exchanges for deciding which type of exchange stage it should suggest. This paper is concerned only with the QI–MDP module. The HMM Module was studied in [13].

4 Personality-Based QI-MDPs

4.1 The States

Consider an h-centered scale of exchange values built as explained in Sect.2. Let $\hat{E}_h = \{E_h^{-n-\frac{sn}{L}}, \ldots, E_h^{-1}, E_h^0, E_h^1, \ldots, E_h^{n-\frac{sn}{L}}\}$ be the set of $2n+1$ equivalence classes of intervals, where, for each value $i = -n - \frac{sn}{L}, \ldots, n - \frac{sn}{L}$:

$$E_h^i = \begin{cases} \{X \mid h + i\frac{L}{n} \leq mid(X) < h + (i+1)\frac{L}{n}\} & \text{if } -n - \frac{hn}{L} \leq i < -1 \\ \{X \mid h - \frac{L}{n} \leq mid(X) < h - \epsilon\} & \text{if } i = -1 \\ \{X \mid h - \epsilon \leq mid(X) \leq h + \epsilon\} & \text{if } i = 0 \\ \{X \mid h + \epsilon < mid(X) \leq h + \frac{L}{n}\} & \text{if } i = 1 \\ \{X \mid h + (i-1)\frac{L}{n} < mid(X) \leq h + i\frac{L}{n}\} & \text{if } 1 < i \leq n - \frac{hn}{L}. \end{cases}$$

[3] We consider systems composed by two agents for simplicity. The results are readily extended for more than two agents using the matrix-like notation introduced in [4], where we assumed that the exchanges performed by any two agents are totally independent and cause no interference in any other exchanges. Thus, subQI–MDPs for any two agents can be solved individually and an optimal global supervisor recommendation realized by concurrent execution of the optimal local recommendations; solution time is determined by the size of the subQI–MDPs [11,12].

The classes $E_h^i \in \hat{E}_h$ are the supervisor representations of classes of material results that are either intervals around the reference value h ($i = 0$), or down scale intervals ($i < 0$), called unfavorable results, or up scale intervals ($i > 0$), called favorable results. Whenever it is understood from the context, we shall denote by E^- (or E^+) any class $E_h^{i<0}$ (or $E_h^{i>0}$). The range of the midpoints of the intervals that belong to a class E_h is called the *representative* of the class E_h. In the following, we identify a class E_h with its representative.

The *states* of a QI–MDP model are pairs $(E_{h_{\alpha\beta}}, E_{h_{\beta\alpha}})$ of equivalence classes representing the material results of the social exchange process between the agents α and β, from the point of view of α and β, respectively, considering their respective $h_{\alpha\beta}$-centered and $h_{\beta\alpha}$-centered scales of exchange values. The set of states is denoted by $\mathbf{E}_{h_{\alpha\beta},h_{\beta\alpha}}$. $(E^0_{h_{\alpha\beta}}, E^0_{h_{\beta\alpha}})$ is the *terminal* state, representing that the system is stable around the reference point $(h_{\alpha\beta}, h_{\beta\alpha})$ that encloses the stability point $(\omega_{\alpha\beta}, \omega_{\beta\alpha})$. In the following, for simplicity, a class $E_{h_{\alpha\beta}}$ is denoted by $E_{\alpha\beta}$, whenever $h_{\alpha\beta}$ is clear from the context.

4.2 The Actions

An *action* is a pair of intervals $(A^i_{\alpha\beta}, A^j_{\beta\alpha})$ that induces a state transition of the form $(E^i_{\alpha\beta}, E^j_{\beta\alpha}) \overset{(A^i_{\alpha\beta}, A^j_{\beta\alpha})}{\mapsto} (E^{i'}_{\alpha\beta}, E^{j'}_{\beta\alpha})$, such that $mid(E^i_{\alpha\beta} + A^i_{\alpha\beta}) \in E^{i'}_{\alpha\beta}$ and $mid(E^j_{\beta\alpha} + A^j_{\beta\alpha}) \in E^{j'}_{\beta\alpha}$. An *interval* $A^i_{\alpha\beta}$ (analogously for an interval $A^j_{\beta\alpha}$) is of one of the following types: (i) a *compensation interval* $C^i_{\alpha\beta}$ of a class representative $E^i_{\alpha\beta}$; (ii) a *go-forward-k-step interval* F^i_k, such that $mid(E^i_{\alpha\beta} + F^i_k) \in E^{(i+k)\neq 0}_{\alpha\beta}$, $i \neq L$; (iii) a *go-backward-k-step interval* B^i_{-k}, such that $mid(E^i_{\alpha\beta} + B^i_{-k}) \in E^{(i-k)\neq 0}_{\alpha\beta}$, $i \neq -L$. The sets of compensation, go-forward and go-backward intervals are denoted by $\mathcal{C}$ (Table 1), $\mathcal{F}$ and $\mathcal{B}$, respectively.

For example, considering a class E^i_h, with $1 < i \leq n - \frac{hn}{L}$, a *go-forward-k-step interval*, with $k \leq n - \frac{hn}{L} - i$, is given by $F^i_k = [k\frac{L}{n} - \epsilon, k\frac{L}{n} + \epsilon]$. And, for a class E^i_h where $-n - \frac{hn}{L} \leq i < -1$, a *go-backward-k-step interval*, with $k \leq n + \frac{hn}{L} + i$, is given by $B^i_{-k} = [-k\frac{L}{n} - \epsilon, -k\frac{L}{n} + \epsilon]$.

Given a target stability point $(\omega_{\alpha\beta}, \omega_{\beta\alpha}) \in \mathbb{R} \times \mathbb{R}$ (which specifies the pair of reference values $(h_{\alpha\beta}, h_{\beta\alpha}) \in H \times H$), the stability supervisor has to find, for each state $(E_{h_{\alpha\beta}}, E_{h_{\beta\alpha}})$ representing the current material results, the action that may achieve the terminal state $(E^0_{h_{\alpha\beta}}, E^0_{h_{\beta\alpha}})$ (representing that the system is stable

Table 1. Specification of compensation intervals

State	Compensation Interval $C^i_h \in \mathcal{C}$
$E_h^{i,-n\leq i<-1}$	$[-(\frac{2i+1}{2}\frac{L}{n}) - \epsilon, -(\frac{2i+1}{2}\frac{L}{n}) + \epsilon]$
E_h^{-1}	$[\frac{1}{2}(\frac{L}{n} + \epsilon) - \epsilon, \frac{1}{2}(\frac{L}{n} + \epsilon) + \epsilon]$
E_h^{0}	$[0, 0]$
E_h^{1}	$[-\frac{1}{2}(\frac{L}{n} + \epsilon) - \epsilon, -\frac{1}{2}(\frac{L}{n} + \epsilon) + \epsilon]$
$E_h^{i,1<i\leq n}$	$[\frac{(1-2i)}{2}\frac{L}{n} - \epsilon, \frac{(1-2i)}{2}\frac{L}{n} + \epsilon]$

around $(h_{\alpha\beta}, h_{\beta\alpha})$) or, at least, another state from where the terminal state can be achieved, with the least number of steps.[4] Such action generates an *optimal exchange recommendation*, consisting of a partially defined exchange stage that the agents are suggested to perform (see Sect. 4.4). This partial definition shall be completed by the analysis of the virtual results, which allows the specification of which particular types of exchange stages (I or II) should be considered.

4.3 Exchanges Between Personality-Based Agents

We define different levels of obedience to the supervisor that the agents may present: *blind obedience* (the agent always follows the recommendations); *eventual obedience* (the agent may not follow the recommendations, according to a certain probability); and *full disregard of recommendations* (the agent always decides on its own, disregarding what was recommended).

The agents may have different personality traits that give rise to different state-transition functions, which specify, for each obedience level, and given the current state and recommendation, a probability distribution $\Pi(\mathbf{E}_{h_{\alpha\beta},h_{\beta\alpha}})$ over the set of states $\mathbf{E}_{h_{\alpha\beta},h_{\beta\alpha}}$ that the interacting agents will try to achieve next. In the following, we illustrate some of those personality traits:

Egoism: the agent is mostly seeking his own benefit, with a high probability to accept exchanges that represent transitions to favorable results;
Strong Egoism: the agent has a very low probability to accept exchanges that represent reduction of its material results even if the agent is maintained in favorable results;
Altruism: the agent is mostly seeking the benefit of the other, with a high probability to accept exchanges that represent transitions toward states where the other agent has favorable results;
Strong Altruism: the agent has a very low probability to accept exchanges that represent reduction of the other agent's material results even if the latter is maintained in favorable results;
Fanaticism: the agent has a very high probability to accept exchanges that lead it to its reference value, avoiding other kinds of transitions;
Tolerance: the agent has a high probability to accept exchanges that lead it to its reference value if his material results are far from that state, but it accepts other kinds of transitions;
Prudence: the agent has a high probability to avoid exchanges when the values involved are higher than a specified limit.

Let $E_h = \{E^-, E^0, E^+\}$ be a simplification of the set $\hat{E}_h$ of the classes of material results, where E^+ and E^- denote the subsets of classes of unfavorable and favorable results, respectively, related to the reference value h. Table 2 presents a pattern of the probability distribution $\Pi(E_h)$, considering individual agent

[4] The choice of actions is constrained by the rules of the social exchanges. Since some transitions are *forbidden* (e.g., both agents increasing results simultaneously), the supervisor has to find alternative paths to lead the agents to the stability point.

Table 2. A pattern of probability distribution $\Pi(E_h)$ for agent transitions

	Egoist agents			Altruist agents		
$\Pi(E_h)$	E^0	E^+	E^-	E^0	E^+	E^-
E^0	low	very-high	very-low	low	very-low	very-high
E^+	low	very-high	very-low	low	very-low	very-high
E^-	low	very-high	very-low	low	very-low	very-high
	Fanatic agents			**Tolerant agents**		
$\Pi(E_h)$	E^0	E^+	E^-	E^0	E^+	E^-
E^0	very-high	very-low	very-low	high	low	low
E^+	very-high	very-low	very-low	high	low	low
E^-	very-high	very-low	very-low	high	low	low

Table 3. A pattern of distribution $\Pi(T)$ for the set T of transitions $E_h^i \mapsto E_h^j$

$\Pi(T)$	$E_h^i < E_h^j$	$E_h^i = E_h^j$	$E_h^i > E_h^j$
Strong Egoism	very-high	low	very-low
Strong Altruism	very-low	low	very-high

transitions, characterizing egoist/altruist and fanatic/tolerant agents. Observe that, for an egoist agent, transitions ending in favorable results (E^+) occurs with very high probability, whereas, for an altruist agent, the most probable transitions are those ending in unfavorable results (E^-). For a fanatic agent, the least probable transitions are those not ending in the terminal state E^0 (around the stability point). In contrast, a tolerant agent accepts transitions to states other than E^0, although with a low probability.

Table 3 shows a pattern for the probability distribution $\Pi(T)$ for the set T of individual agent transitions $E_h^i \mapsto E_h^j$, for strong egoism/altruism. Observe that strong egoist agents presents a very high probability to increase their material results in any exchange, whereas strong altruist agents behave in a completely opposite way.

Table 4 shows parts of sample state-transition functions **F** for systems composed by (a) two tolerant agents and (b) two egoist agents that always disregard the supervisor's recommendations. The mark **X** indicates that the transition is forbidden according to the adopted social rules (both agents increasing results simultaneously). In (b), the highest probabilities appear in the transitions ending in the state (E^+, E^+), representing increasing results for both agents, or in the states $(-, E^+)$ or $(E^+, -)$ when the transitions to the state (E^+, E^+) are not allowed. The probability of 100% in the last line of (b) indicates that the agents refuse to exchange (which would lead both to unfavorable results), remaining in the same state (E^-, E^-). This shows that this system presents an absorbent state, (E^-, E^-), meaning that the system is not able to leave that state if it reaches it, and so it may never achieve the desired target stability point. In

Table 4. Parts of state-transition functions **F** for pairs of agents that always disregard recommendations

(a) (tolerant, tolerant) agents

F (%)	(E^0,E^0)	(E^0,E^+)	(E^0,E^-)	(E^+,E^0)	(E^+,E^+)	(E^+,E^-)	(E^-,E^0)	(E^-,E^+)	(E^-,E^-)
(E^0,E^0)	63.90	**X**	13.70	**X**	**X**	2.90	13.70	2.90	2.90
(E^+,E^-)	49.20	10.50	10.50	10.50	2.20	2.20	10.50	2.20	2.20
(E^-,E^-)	**X**	**X**	37.85	**X**	**X**	8.10	37.85	8.10	8.10

(b) (egoist, egoist) agents

F (%)	(E^0,E^0)	(E^0,E^+)	(E^0,E^-)	(E^+,E^0)	(E^+,E^+)	(E^+,E^-)	(E^-,E^0)	(E^-,E^+)	(E^-,E^-)
(E^0,E^-)	**X**	**X**	0.00	**X**	**X**	0.00	15.00	85.00	0.00
(E^+,E^+)	2.20	12.00	0.70	12.00	64.10	4.00	0.70	4.00	0.30
(E^+,E^-)	2.20	12.80	0.00	12.00	68.00	0.00	0.70	4.30	0.00
(E^-,E^-)	**X**	**X**	0.00	**X**	**X**	0.00	0.00	0.00	100.00

Table 5. Parts of state-transition functions **F** for pair of agents with 50% of obedience

(a) (tolerant, tolerant) agents

F (%)	(E^0,E^0)	(E^0,E^+)	(E^0,E^-)	(E^+,E^0)	(E^+,E^+)	(E^+,E^-)	(E^-,E^0)	(E^-,E^+)	(E^-,E^-)
(E^0,E^0)	81.95	**X**	6.85	**X**	**X**	1.45	6.85	1.45	1.45
(E^+,E^-)	74.6	5.25	5.25	5.25	1.10	1.10	5.25	1.10	1.10
(E^-,E^-)	**X**	**X**	18.92	**X**	**X**	29.05	18.92	29.05	4.06

(b) (egoist, egoist) agents

F (%)	(E^0,E^0)	(E^0,E^+)	(E^0,E^-)	(E^+,E^0)	(E^+,E^+)	(E^+,E^-)	(E^-,E^0)	(E^-,E^+)	(E^-,E^-)
(E^0,E^-)	**X**	**X**	0.0%	**X**	**X**	25.00	7.50	67.50	0.00
(E^+,E^+)	51.10	6.00	0.35	6.00	32.05	2.00	0.35	2.00	0.15
(E^+,E^-)	51.10	6.40	0.00	6.00	34.00	0.00	0.35	2.15	0.00
(E^-,E^-)	**X**	**X**	0.00	**X**	**X**	25.00	0.00	25.00	50.00

(a), one observes the more uniform behavior of tolerant agents, even though the transitions to the states (E^0,E^0), $(E^0,-)$ and $(-,E^0)$ being the most probable.

We remark that even if the agents present a certain level of obedience, there may be a great deal of uncertainty about the effects of the supervisor's recommendations. Considering an obedience level of 50%, the state-transition functions shown in Table 4 become the respective ones shown in Table 5, showing an increase in the probability of the transitions ending in (E^0,E^0) and also the absence of an absorbent state.

For example, for two agents α and β and classes of material results given by $(E^i_{h_{\alpha\beta}}, E^j_{h_{\beta\alpha}}) \equiv ([h_{\alpha\beta} + i\frac{L}{n}, h_{\alpha\beta} + (i+1)\frac{L}{n}], [h_{\beta\alpha} + (j-1)\frac{L}{n}, h_{\beta\alpha} + j\frac{L}{n}])$, with $-n - \frac{nh_{\alpha\beta}}{L} \leq i < -1$ and $1 < j \leq n - \frac{nh_{\beta\alpha}}{L}$, a *compensation–compensation* action $(C^i_{h_{\beta\alpha}}, C^j_{h_{\beta\alpha}}) \equiv ([-\frac{2i+1}{2}\frac{L}{n} - \epsilon, -\frac{2i+1}{2}\frac{L}{n} + \epsilon], [\frac{(1-2j)}{2}\frac{L}{n} - \epsilon, \frac{(1-2j)}{2}\frac{L}{n} + \epsilon])$, should be chosen by the stability supervisor; then, if the agents are obedient, and under certain conditions (see Sect. 4.5), the resulting state transition would be one of the following, with $-n - \frac{nh_{\alpha\beta}}{L} \leq i < -1$ and $1 < j \leq n - \frac{nh_{\beta\alpha}}{L}$: $(E^i_{h_{\alpha\beta}}, E^j_{h_{\beta\alpha}}) \mapsto (E^0_{h_{\alpha\beta}}, E^0_{h_{\beta\alpha}})$ or $(E^{-1}_{h_{\alpha\beta}}, E^0_{h_{\beta\alpha}})$ or $(E^0_{h_{\alpha\beta}}, E^1_{h_{\beta\alpha}})$ or $(E^{-1}_{h_{\alpha\beta}}, E^1_{h_{\beta\alpha}})$, in increasing order of probability. If one of the agents is not obedient, then there is a probability that none of the above transitions occurs.

4.4 Optimal Value Recommendations

A reward function $\mathbf{R} : (\mathbf{E} \times \mathbf{A}) \to \mathbb{R}$ must conform to the idea of supporting a recommendation function that is able to direct pairs of agents into the stability point, according to the model of social exchanges (see, e.g, [7]). One sample reward function is partially sketched in Table 6, illustrating some requirements that should be satisfied by such functions. For instance, if the current state is of the type (E^-, E^+), then the reward function must state that the best action to be chosen is a *compensation-compensation* action (C, C), which may result in a state transition $(E^-, E^+) \mapsto (E^0, E^0)$. On the other hand, if the current state is of type (E^-, E^-), then it must prevent the choice of a *compensation-compensation* action (C, C) that would generate a recommendation of exchange of *satisfaction-satisfaction* type, which is forbidden in the model, because it considers impossible to get a satisfaction from no service.

Table 6. Partial schema of the reward function R

R	(C,C)	(B_{-1},F_{+1})	(F_{+1},B_{-1})	(B_{-3},C)
(E^-,E^+)	30	-5	3	20
(E^+,E^+)	30	0	0	20
(E^-,E^-)	-30	30	30	26

The *optimal recommendation* associated to an optimal policy π^* is an operator ρ_{π^*} that gives, for each state $(E^i_{\alpha\beta}, E^j_{\beta\alpha})$ and optimal action $\pi^*(E^i_{\alpha\beta}, E^j_{\beta\alpha}) = (A^i_{\alpha\beta}, A^j_{\beta\alpha})$, partial definitions of recommended exchange stages, consisted by either $(r_{\alpha\beta}, A^i_{\alpha\beta})$ and $(s_{\beta\alpha}, A^j_{\beta\alpha})$, or $(s_{\alpha\beta}, A^i_{\alpha\beta})$ and $(r_{\beta\alpha}, A^j_{\beta\alpha})$, where $(r_{\lambda\delta}, W)$ means the realization, by the agent λ, of a service with investment value W, and $(s_{\delta\lambda}, W')$ means δ's satisfaction with value W', for receiving the service. The optimal recommendation ρ_{π^*} is partially sketched in Table 7.

Finally, the stability supervisor has to decide which types of exchange stages (I or II) should be recommended. This is done by the analysis of the virtual results. For example, if $\mathbf{v}_{\alpha\beta} > 0$ ($\mathbf{v}_{\beta\alpha} > 0$), then α (β) is able to charge β (α) the credit for services previously done. In this case, an exchange stage T^1 (T^2) of type $\mathrm{II}_{\alpha\beta}$ ($\mathrm{II}_{\beta\alpha}$) should be recommended. However, if $\mathbf{v}_{\alpha\beta} \leq 0$ ($\mathbf{v}_{\beta\alpha} \leq 0$), then the agent α (β) does not have any credit to charge α (β). Therefore, the service done by the agent β (α) must be spontaneous. In this case, an exchange stage T^3 (T^4) of type $\mathrm{I}_{\beta\alpha}$ ($\mathrm{I}_{\alpha\beta}$) should then be recommended. Some stage recommendations and their combined effects with the optimal value recommendations are sketched in the simplified state transition diagram shown in Fig. 2, where the dot lines represent alternative paths that were not considered as optimal recommendations since they may seem unfair according to social rules.

4.5 Formal Definition and Analysis of the Stabilization Process

Definition 1. *A* Qualitative Interval Markov Decision Process (QI–MDP)*, for keeping the social exchanges in a multiagent system stable around a reference value h, is a tuple* $\langle \mathbf{E}_{h_{\alpha\beta},h_{\beta\alpha}}, \mathbf{A}, \mathbf{F}, \mathbf{R}\rangle^{L,n}_{\epsilon}$*, where:*

Table 7. Partial schema of the optimal value recommendation ρ_{π^*}

State	Optimal policy	Recommendation	Label
$(E^i,E^j)^{1<j\leq n}_{-n\leq i<-1}$	(C^i,C^j)	$((r_{\beta\alpha},C^j),(s_{\alpha\beta},C^i))$	R_1
$(E^i,E^j)_{1<i,j\leq n}$	(C^i,C^j)	$((r_{\alpha\beta},C^i),(s_{\beta\alpha},C^j))$ or $((r_{\beta\alpha},C^j),(s_{\alpha\beta},C^i))$	R_2 R_3
$(E^0,E^j)_{1<j\leq n}$	$([0,0],C^j)$	$((r_{\beta\alpha},C^j),(s_{\alpha\beta},[0,0]))$	R_4
$(E^0,E^i)_{-n\leq i<-1}$	$(B^0_{-1},F^i_{+(-i+1)})$	$((r_{\alpha\beta},B^0_{-1}),(s_{\beta\alpha},F^i_{+(-i+1)}))$	R_5
$(E^{-1},E^j)_{1<j\leq n}$	$(F^{-1}_{+1}\vee C^{-1},C^j)$	$((r_{\beta\alpha},C^j),(s_{\alpha\beta},F^{-1}_{+1}$ or $C^{-1}))$	R_6
$(E^1,E^i)_{-n\leq i<-1}$	$(B^1_{-1}\vee C^1,C^i)$	$((r_{\alpha\beta},B^1_{-1}\vee C^1),(s_{\beta\alpha},C^i))$	R_7
(E^{-1},E^1)	$(F^{-1}_{+1}\vee C^{-1},B^1_{-1}\vee C^1)$	$((r_{\beta\alpha},B^1_{-1}\vee C^1),(s_{\alpha\beta},F^{-1}_{+1}\vee C^{-1}))$	R_8
(E^1,E^{-1})	$(B^1_{-1}\vee C^1,F^{-1}_{+1}\vee C^{-1})$	$((r_{\alpha\beta},B^1_{-1}\vee C^1),(s_{\beta\alpha},F^{-1}_{+1}\vee C^{-1}))$	R_9
$(E^i,E^1)_{-n\leq i<-1}$	$(C^i,B^1_{-1}\vee C^1)$	$((r_{\beta\alpha},B^1_{-1}\vee C^1),(s_{\alpha\beta},C^i))$	R_{10}
(E^{-1},E^0)	$(F^{-1}_{+1}\vee C^{-1},B^0_{-1})$	$((r_{\beta\alpha},B^0_{-1}),(s_{\alpha\beta},F^{-1}_{+1}\vee C^{-1}))$	R_{11}
(E^0,E^{-1})	$(B^0_{-1},F^{-1}_{+1}\vee C^{-1})$	$((r_{\alpha\beta},B^0_{-1}),(s_{\beta\alpha},F^{-1}_{+1}\vee C^{-1}))$	R_{12}
$(E^i,E^j)_{-n\leq i,j<-1}$	$(F^i_{+(-i+1)},B^j_{-1})$ or $(B^j_{-1},F^i_{+(-i+1)})$	$((r_{\beta\alpha},B^j_{-1}),(s_{\alpha\beta},F^i_{+(-i+1)})$ or $((r_{\alpha\beta},B^j_{-1}),(s_{\beta\alpha},F^i_{+(-i+1)})$	R_{13} R_{14}

- *The set of states of the model is* $\mathbf{E}_{h_{\alpha\beta},h_{\beta\alpha}}=\{(E^i_{\alpha\beta},E^j_{\beta\alpha})\mid E^i_{\alpha\beta}\in\hat{E}_{h_{\alpha\beta}},E^j_{\beta\alpha}\in\hat{E}_{h_{\beta\alpha}}\}$ *of pairs of classes of material results as specified in Sect. 4.1.*
- *The set of the actions of the model is the set* $\mathbf{A}=\{(A^i_{\alpha\beta},A^j_{\beta\alpha})\mid A^i_{\alpha\beta},A^j_{\beta\alpha}\in\mathcal{C}\cup\mathcal{F}\cup\mathcal{B}\}$ *of pairs of compensation, go-forward and go-backward intervals.*
- $\mathbf{F}:\mathbf{E}_{h_{\alpha\beta},h_{\beta\alpha}}\times\mathbf{A}\rightarrow\Pi(\mathbf{E}_{h_{\alpha\beta},h_{\beta\alpha}})$ *is the personality-based state-transition function, that gives for each state and each action, a probability distribution over the set of states* $\mathbf{E}_{h_{\alpha\beta},h_{\beta\alpha}}$.
- $\mathbf{R}:\mathbf{E}_{h_{\alpha\beta},h_{\beta\alpha}}\times\mathbf{A}\rightarrow\mathbb{R}$ *is the reward function, giving the expected reward gained by choosing an action* $(A^i_{\alpha\beta},A^j_{\beta\alpha})$ *when the current state is* $(E^i_{\alpha\beta},E^j_{\beta\alpha})$.

The analysis of the stabilization process is concerned with the number of steps that are necessary to achieve the target stability point. Since the decision process is non-trivial (due to: (i) the qualitative nature of exchange values, (ii) the restrictions imposed by the definition of exchange, that always requires a service to be done in any stage, and mainly (iii) the stochastic nature of the agents' behaviors), an analytical study was only possible for agents with blind obedience (after a certain number of free exchanges). Then, the supervisor accuracy $\kappa_n=\frac{L}{n}$ can be adjusted to have the system stable in *at most four steps*, as we show here.[5] Let $\mathbf{m}^\tau_{\alpha\beta}$ and $\mathbf{m}^\tau_{\beta\alpha}$ be the material results of an exchange process performed by the agents α and β, at step τ, and $h_{\alpha\beta}$ and $h_{\beta\alpha}$ be the reference values that approximate a target stability point. For a tolerance ϵ, it holds that:

Proposition 1. *If* $\mathbf{m}^0_{\alpha\beta}\in E^{-1}_{h_{\alpha\beta}}$ *and* $\mathbf{m}^0_{\beta\alpha}\in E^1_{h_{\beta\alpha}}$, *then the target stability point is achieved in one step if and only if* $1<\frac{\kappa_n}{\epsilon}\leq 3$.

Proof. ($\Rightarrow$) Since $h_{\beta\alpha}+\epsilon<mid(\mathbf{m}^0_{\beta\alpha})\leq h_{\beta\alpha}+\frac{L}{n}$ and the optimal recommendation (Table 7, $R8$) gives the action $C^1_{h_{\beta\alpha}}=[-\frac{1}{2}(\frac{L}{n}+\epsilon),-\frac{1}{2}(\frac{L}{n}+\epsilon)]$ (Table 1), it

[5] For other levels of obedience, the analysis is based on simulations, as shown in Sect. 5.

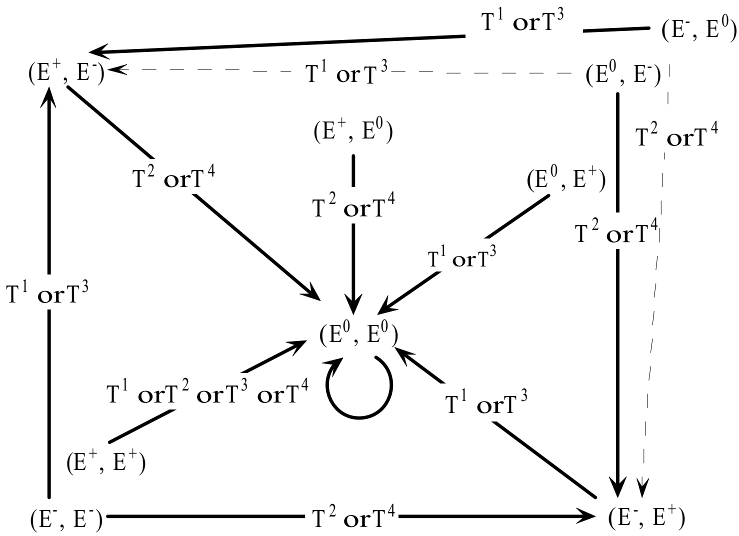

Fig. 2. Effects of stage and optimal value recommendations

follows that: $h_{\beta\alpha}+\epsilon-\frac{1}{2}(\frac{L}{n}+\epsilon) < mid(\mathbf{m}^0_{\beta\alpha})-\frac{1}{2}(\frac{L}{n}+\epsilon) \leq h_{\beta\alpha}+\frac{L}{n}-\frac{1}{2}(\frac{L}{n}+\epsilon) \Rightarrow h_{\beta\alpha}+\frac{1}{2}(-q\epsilon+\epsilon) < mid(\mathbf{m}^1_{\beta\alpha}) \leq h_{\beta\alpha}+\frac{1}{2}(q\epsilon-\epsilon)$, where $\frac{L}{n}=q\epsilon$, $q>1$. If the system achieves the stability point in the step 1, then $h_{\beta\alpha}+\frac{1}{2}(q\epsilon-\epsilon) \leq h_{\beta\alpha}+\epsilon$, $1<q\leq 3$, and thus, $1<\frac{\kappa_n}{\epsilon}\leq 3$, since $\kappa_n=\frac{L}{n}$.

Proposition 2. *(i) If* $\mathbf{m}^0_{\alpha\beta} \in E^i_{h_{\alpha\beta}}$, $1<i\leq n$, *then it is possible to get* $\mathbf{m}^\tau_{\alpha\beta} \in E^0_{h_{\alpha\beta}}$ *in at most* $\tau=2$ *steps if and only if* $1<\frac{\kappa_n}{\epsilon}\leq 3$*; (ii) If* $\mathbf{m}^0_{\beta\alpha} \in E^i_{h_{\beta\alpha}}$, $-n\leq i<-1$, *then it is possible to get* $\mathbf{m}^\tau_{\beta\alpha} \in E^0_{h_{\beta\alpha}}$ *in at most* $\tau=2$ *steps if and only if* $1<\frac{\kappa_n}{\epsilon}\leq 3$*; (iii) If* $\mathbf{m}^0_{\alpha\beta} \in E^i_{h_{\alpha\beta}}$, *with* $1<i\leq n$ *and* $h_{\alpha\beta}+\frac{2i+1}{2}\frac{L}{n}-\epsilon \leq mid(\mathbf{m}^0_{\alpha\beta}) \leq h_{\alpha\beta}+\frac{2i+1}{2}\frac{L}{n}+\epsilon$, *then* $\mathbf{m}^1_{\alpha\beta} \in E^0_{h_{\alpha\beta}}$.

Proof. (i)($\Rightarrow$) Since $h_{\alpha\beta}+(i-1)\frac{L}{n} \leq mid(\mathbf{m}^0_{\alpha\beta}) < h_{\alpha\beta}+i\frac{L}{n}$ and the optimal recommendation (Table 7, $R2/R3$) is the action $C^i=[\frac{(1-2i)}{2}\frac{L}{n},\frac{(1-2i)}{2}\frac{L}{n}]$ (Table 1), it follows that: $h_{\alpha\beta}+(i-1)\frac{L}{n}+\frac{(1-2i)}{2}\frac{L}{n} < mid(\mathbf{m}^0_{\beta\alpha})+\frac{(1-2i)}{2}\frac{L}{n} \leq h_{\alpha\beta}+i\frac{L}{n}+\frac{(1-2i)}{2}\frac{L}{n} \Rightarrow h_{\alpha\beta}-\frac{1}{2}\frac{L}{n} < mid(\mathbf{m}^1_{\beta\alpha}) \leq h_{\alpha\beta}+\frac{1}{2}\frac{L}{n}$, and thus $\mathbf{m}^1_{\beta\alpha} \in E^1_{\alpha}$. From Prop. 1, with one more step we can get the result.

It follows that an individual transition from E^i ($1<i\leq n$ or $-n\leq i<-1$), to the stability point can be done in at most two steps ($E^i \mapsto E^1$(or E^{-1}) $\mapsto E^0$). However, combined transitions departing from a state (E^i,E^j) or (E^j,E^i), with $1<i\leq n$ and $-n\leq j<-1$, may result in a state different from (E^1,E^{-1}), (E^{-1},E^1) or (E^0,E^0). The worst case is when the system is in the state (E^i,E^j), with $-n\leq i,j<-1$, since two simultaneous positive compensation actions are not allowed. In this case, which occurs very often in exchanges between altruist agents, the optimal recommendation (Table 7) leads the agents to the stability point in at most four steps, by one of the transitions: (i) $(E^i, E^j)_{-n\leq i,j<-1} \overset{R13}{\mapsto} (E^1, E^j)_{-n\leq j<-1} \overset{R7}{\mapsto} (E^0, E^{-1}) \overset{R12}{\mapsto} (E^{-1}, E^1) \overset{R8}{\mapsto} (E^0, E^0)$ or (ii) $(E^i, E^j)_{-n\leq i,j<-1} \overset{R14}{\mapsto} (E^j, E^1)_{-n\leq j<-1} \overset{R10}{\mapsto} (E^{-1}, E^0) \overset{R11}{\mapsto} (E^1, E^{-1}) \overset{R9}{\mapsto} (E^0, E^0)$.

5 A Sample Simulation

We show a simulation of part of the scenario analyzed in [14], extending the application to consider personality-based agents. The situation is a political one, with politicians and voters interacting for the purpose of electing politicians to governmental positions. Politicians are expected to fulfill the promises they have made to voters before the election, by making decisions that favor the voter's interests. After reaching governmental positions, politicians may or may not fulfill their promises. In the positive case, they are entitled to charge the voters for their re-election in the next polling. On the other hand, voters are expected to choose politicians that best represent their interests, and give them votes. After the election, they are entitled to charge the politicians for coherent behavior with the promises they made. Frustration of any of those expectations entitles the frustrated agent to refrain from behaving in a positive way toward the others.

An equilibrated political society is one where both voters and politicians do not accumulate neither benefits nor losses, which is an idealization that may never occur in practice. On the other hand, a stable political society is one where both voters and politicians behave as respectively expected by the others during a considered period, or the regulation of the behaviors of politicians and voters is such that significant deviations from the expected behaviors of any of them get each of the agents to be either enforced to backtrack from the deviated behavior or allowed to look for other partners with different interests. In our simulations, such regulation actions are not allowed, so that agents are doomed to misfortune if the stability supervisor fails in being effective.

Exchange values can easily be associated with each action, of voting and governmental decision. Voters and politicians can thus successively build a balance of such values, as elections are successively performed. Considering this as an open society, at each election new voters and new politicians may appear in the process, behaving as non-transparent agents for the supervisor.

In a realistic simulation, both politicians and voters would have their own decision procedures about the actions they have to take at each election. Such procedures can be seen as stability supervisors that were internalized at each agent, and that restrict themselves to give recommendations specifically for the agent where each one is internalized. Having been internalized, the supervisors can easily be seen to operate under the condition of partial observation, since the internalization makes it not possible for the agents to fully grasp the exchange values accumulated by the others. We are leaving for future work the problem of tackling such situations, also because it could involve the analysis of interactions between groups of agents, where the results of the exchanges between a pair of agents may influence the exchanges performed by the others.

Here, the stability supervisor is a centralized agent that makes recommendations for a pair of transparent agents: a politician and an elector. The simulations were developed in the Python programming language, generating (i) tables with the configurations of exchange values and material results at each time $t \in T = 0, \ldots, 1000$, and (ii) graphics showing the trajectory of the mid points of the material results of the exchanges. The material and virtual values that the electors and politicians could use at each exchange were set to vary in $-100 \ldots + 100$. The target stability point was set at $\omega = 1000$ for both agents, meaning that both agents get positive material results from the interactions. A tolerance of $\epsilon = 25$ was adopted for the definition of the reference value.

At each election and successive governmental period, the elector β assigns a value $r_{\beta\alpha}$ to its vote for the politician α and concludes a value $v_{\beta\alpha}$ for his credit over his election. Correspondingly, the politician α assigns a value $r_{\alpha\beta}$ to the decisions he makes while in the government, after the election, and a credit $v_{\alpha\beta}$ for the benefits he thinks are received by the elector β due to those decisions. Satisfaction and debit values $(s_{\alpha\beta}, s_{\beta\alpha}, t_{\alpha\beta}, t_{\beta\alpha})$ are correspondingly assigned.

First, we considered successions of elections where the elector and the politician *always disregard the supervisor's recommendations*. In those simulations (Fig. 3(a), exchange values bound to $[-3000, 2400]$), the exchanges were totally guided by the agents' personality traits, characterized by the *egoism* of the politician and the *altruism* of the elector. The politician profited from the interaction much more than the elector, which kept the latter in unfavorable results (related to the stability point), resulting that the system was unable to be stabilized.

In successive experiments, we increased the level of obedience to the recommendations, generating the following simulations: (1) obedience during 2% of the elections (Fig.3(b), with exchange values bound to $[-850, 1800]$ and range of exchange values equals to 2650); (2) obedience during 25% of the elections (Fig. 3(c), exchange values bound to $[500, 1500]$, maximal deviation of 500 around the stability point); (3) obedience in 100% of the elections (Fig. 3(d), exchange values bound to $[900, 1100]$, maximal deviation of 100 around the stability point).

Figure 3(b) shows a succession of elections, with a level of obedience to the recommendations of 2%. Observe that just such level of obedience was enough to make the politician and the elector alternated their kinds of behaviors, thus avoiding that one of them profited from the interaction at the expense of the other. The system was able to pass through the stability point in various opportunities (e.g., at $t = 180$ and $t = 365$), but was unstable almost all of the time. Figure 3(c-d) shows the effects of the increasing level of obedience. The range of deviations of the results from the stability point was progressively reduced as the politician and the elector progressively adhered to the recommendations.

The simulations that we produced seem to agree with the theoretical predictions derived from the model (in Sect. 4.5). Thus, considering agents with *blind obedience* (Fig. 3(d)), the deviation around the stability point stayed stable between -100 and 100, the maximum variation allowed at each interaction.

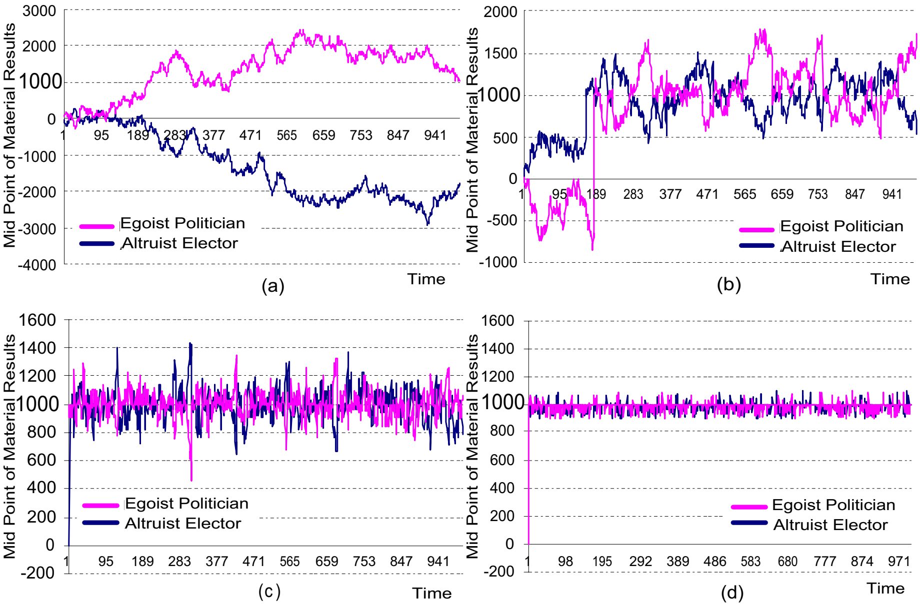

Fig. 3. (a) Agents always disregarding recommendations; agent obedience in (b) 2% of the elections, (c) 25% of the elections and (d) 100% of the elections (blind obedience)

6 Related Work

Values have been extensively used in the MAS area, through value-based and market-oriented decision, and value-based social theory (see, e.g., [15,16,17]), as well social norms (e.g., in [18,19]), the latter considering that knowledge sharing between agents is supported by social contracts and rules.

However, the approach based on social exchange values, which gives rise to a qualitative analysis of the interactions based on the individual evaluations of the exchanges, appeared only in 2003 [20], its formulation becoming stable only after [4]. Since them, the merits in using Piaget's notion of exchange values to the analysis of social organization, and applications to problems like that of partners selection, formation of coalitions and collaborative interactions have been discussed [14,21,22]. In particular, the application of this approach applied to the analysis of successful/uncessful cooperative interactions in the bioinformatics domain was presented in [22].

On the other hand, the study of personality-based agents can be traced back to at least [23], while its study in the context of multiagent systems goes back to [8,24], where advantages and possible applications of the approach were extensively discussed. In both works, personality traits were mapped into goals and practical reasoning rules (internal point of view). Modeling personality traits from an external (the supervisor's) point of view, through state transition matrices as we do here, seems to be new.

7 Conclusion

The paper leads toward the idea of modeling agents' personality traits in social exchange regulation mechanisms in open societies, also extending the previously proposed concept of equilibrium supervisors to consider the stability of social exchanges in points that may be different from the equilibrium and may vary in time. Then, the notion of equilibrium is a particular case of stability [4].

We studied two sample sets of personality traits: (i) blind, eventual obedience, and full disregard of recommendations (related to the levels of adherence to the regulation mechanism), and (ii) fanaticism, tolerance, egoism, altruism and prudence (in connection to preferences about balances of material results).

The regulation mechanism implements a Qualitative Interval Markov Decision Processes for the coordination of the exchanges between transparent agents (agents that allow full external access to their internal states). A theoretical analysis of the Qualitative Interval Markov Decision Process was realized, and simulations of performances of a stability supervisor were presented, considering different levels of obedience, which conformed to the theoretical analysis.

We point out that the regulation of the interactions between non-transparent agents (agents that restrict such external access) was done in [13], with the help of personality-based Hidden Markov Models, so that the supervisor is able to recognize and maintain an adequate model of the personality traits of such agents, based on observations of their behaviors. In that work, the set of personality traits was enlarged with the agents' tendencies in the evaluation of their virtual results, which is then observed by the supervisor.

In our future work, we expect to advance the internalization of the stability supervisor into the agents themselves, going toward the idea of self-regulation of exchange processes, not only distributing the decision process (like, e.g., in Multiagent MDP [11,25]), but also considering incomplete information about the balances of material results of the exchanges between non-transparent agents, in the form of a *personality-based qualitative interval* Partially Observable MDP (POMDP) [26,27], a kind of decentralized or distributed POMDP [28,29,30].

Acknowledgements. This work has been partially supported by FAPERGS and CNPq. We would like to thank the referees and the participants of the COIN@ECAI 2006 workshop for their valuable comments.

References

1. Castelfranchi, C.: Engineering social order. In: Omicini, A., Tolksdorf, R., Zambonelli, F. (eds.) Engineer. Societ. in Agents World, pp. 1–18. Springer, Berlin (2000)
2. Homans, G.C.: The Human Group. Harcourt, Brace & World, New York (1950)
3. Dignum, V., Dignum, F.: Coordinating tasks in agent organizations or: Can we ask you to read this paper? LNCS, vol. 4386, pp. 30–45. Springer, Heidelberg (2007)

4. Pereira Dimuro, G., da Rocha Costa, A.C., Palazzo, L.A.M.: Systems of exchange values as tools for multi-agent organizations (Special Issue on Agents' Organizations). Journal of the Brazilian Computer Society 11(1), 31–50 (2005)
5. Piaget, J.: Sociological Studies. Routlege, London (1995)
6. Moore, R.E.: Methods and Applic. of Interval Analysis. SIAM, Philadelphia (1979)
7. Pereira Dimuro, G., da Rocha Costa, A.C.: Exchange values and self-regulation of exchanges in multi-agent systems: the provisory, centralized model. In: Brueckner, S.A., Serugendo, G.D.M., Hales, D., Zambonelli, F. (eds.) ESOA 2005. LNCS (LNAI), vol. 3910, pp. 75–89. Springer, Heidelberg (2006)
8. Castelfranchi, C., Rosis, F., Falcone, R., Pizzutilo, S.: Personality traits and social attitudes in multiagent cooperation. Applied Artif. Intelligence 12, 649–675 (1998)
9. Pereira Dimuro, G., da Rocha Costa, A.C.: Interval-based Markov Decision Processes for regulating interactions between two agents in multiagent systems. In: Dongarra, J.J., Madsen, K., Waśniewski, J. (eds.) PARA 2004. LNCS, vol. 3732, pp. 102–111. Springer, Heidelberg (2006)
10. Rabiner, L.R.: A tutorial on Hidden Markov Models and selected applications in speech recognition. Proc. of the IEEE 77(2), 257–286 (1989)
11. Boutilier, C.: Sequential optimality and coordination in multiagent systems. In: Proc. XVI Intl. Joint Conf. Artificial Intellig. IJCAI'99, Stockholm, pp. 478–485 (1999)
12. Puterman, M.L.: Markov Decision Processes. Wiley, Chichester, UK (1994)
13. Pereira Dimuro, G., da Rocha Costa, A.C., Vargas Gonçalves, L., Hübner, A.: Regulating social exchanges between personality-based non-transparent agents. In: Gelbukh, A., Reyes-Garcia, C.A. (eds.) MICAI 2006. LNCS (LNAI), vol. 4293, Springer, Heidelberg (2006)
14. Rodrigues, M.R., da Rocha Costa, A.C.: Using qualitative exchange values to improve the modelling of social interactions. In: Multi-Agent-Based Simulation III. LNCS, vol. 2927, pp. 57–72. Springer, Heidelberg (2003)
15. Antunes, L., Coelho, H.: Decisions based upon multiple values: the BVG agent architecture. In: Barahona, P., Alferes, J.J. (eds.) EPIA 1999. LNCS (LNAI), vol. 1695, pp. 297–311. Springer, Heidelberg (1999)
16. Miceli, M., Castelfranchi, C.: The role of evaluation in cognition and social interaction. In: Dautenhahn, K. (ed.) Human cognition and agent technology, pp. 225–262. John Benjamins, Amsterdam (2000)
17. Walsh, W.E., Wellman, M.P.: A market protocol for distributed task allocation. In: Proc. III Intl. Conf. on Multiagent Systems, Paris pp. 325–332 (1998)
18. Castelfranchi, C., Dignum, F., Jonker, C., Treur, J.: Deliberate normative agents: Principles and architecture. In: Jennings, N.R. (ed.) ATAL 1999. LNCS, vol. 1757, pp. 364–378. Springer, Heidelberg (2000)
19. López y López, F., Luck, M., d'Inverno, M.: Constraining autonomy through norms. In: Alonso, E., Kudenko, D., Kazakov, D. (eds.) Adaptive Agents and Multi-Agent Systems. LNCS (LNAI), vol. 2636, pp. 674–681. Springer, Heidelberg (2003)
20. Rodrigues, M.R., da Rocha Costa, A.C., Bordini, R.: A system of exchange values to support social interactions in artificial societetes. In: Proc. II Intl Conf. on Autonomous Agents and Multiag. Systems, AAMAS'03, Melbourne, pp. 81–88. ACM Press, New York (2003)
21. Rodrigues, M.R., Luck, M.: Analysing partner selection through exchange values. In: Sichman, J.S., Antunes, L. (eds.) MABS 2005. LNCS (LNAI), vol. 3891, pp. 24–40. Springer, Heidelberg (2006)
22. Rodrigues, M.R., Luck, M.: Cooperative interactions: an exchange values model. LNCS, vol. 4386, pp. 344–360. Springer, Heidelberg (2007)

23. Carbonell, J.G.: Towards a process model of human personality traits. Artificial Intelligence 15(1-2), 49–74 (1980)
24. Castelfranchi, C., Rosis, F., Falcone, R., Pizzutilo, S.: A testbed for investigating personality-based multiagent cooperation. In: Proc. of the Symp. on Logical Approaches to Agent Modeling and Design, Aix-en-Provence (1997)
25. Boutilier, C.: Multiagent systems: challenges and oportunities for decision theoretic planning. Artificial Intelligence Magazine 20(4), 35–43 (1999)
26. Kaelbling, L.P., Littman, M.L., Cassandra, A.R.: Planning and acting in partially observable stochastic domains. Artificial Intelligence 101(1), 99–134 (1998)
27. Lovejoy, W.S.: A survey of algorithmic methods for Partially Observable Markov Decision Processes. Annals of Operations Research 28(1), 47–65 (1991)
28. Bernstein, D., Zilberstein, S., Immerman, N.: The complexity of decentralized control of Markov Decision Processes. In: Proc. of the 16th Conference on Uncertainty in Artificial Intelligence, Stanford, pp. 32–37 (2000)
29. Nair, R., Tambe, M., Yokoo, M., Pynadath, D., Marsella, S.: Taming decentralized POMDPs: Towards efficient policy computation for multiagent settings. In: Proc. 18th Intl. Joint Conf. on Artificial Intelligence, IJCAI'03, Acapulco, pp. 705–711 (2003)
30. Scerri, P., Pynadath, D., Tambe, M.: Towards adjustable autonomy for the real-world. Journal on Artificial Intelligence Research 17, 171–228 (2002)