SOCRATES: from SOCial Relation to ATtitude ExpressionS

4 authors:

Florian Pecune
Télécom ParisTech
14 PUBLICATIONS   **7** CITATIONS

SEE PROFILE

Magalie Ochs
LSIS
76 PUBLICATIONS   **376** CITATIONS

SEE PROFILE

Stacy Marsella
Northeastern University
221 PUBLICATIONS   **5,936** CITATIONS

SEE PROFILE

Catherine Pelachaud
French National Centre for Scientific Research
360 PUBLICATIONS   **5,850** CITATIONS

SEE PROFILE

SOCRATES: from SOCial Relation to ATtitude ExpressionS

Florian Pecune
LTCI, CNRS, Telecom
ParisTech, Universite
Paris-Saclay
pecune@enst.fr

Magalie Ochs
Aix Marseille Universite,
CNRS, ENSAM, Universite de
Toulon, LSIS UMR7296
13397, Marseille, France
magalie.ochs@lsis.org

Stacy Marsella
Northeastern University
Boston, MA 02115, USA
s.marsella@neu.edu

Catherine Pelachaud
LTCI, CNRS, Telecom
ParisTech, Universite
Paris-Saclay
pelachaud@enst.fr

ABSTRACT

To be perceived as believable partners in human-machine interactions, virtual agents have to express adequate social attitudes. The social attitude expressed by an agent should reflect the social situation of the interaction. The agent ought to take into account its role and its social relation toward its interactants when deciding how to react in the interaction. To build such an agent able to reason about its role and relation and to adapt its social attitude, we build a model of social decision making. First, we formalize the dynamics of the social relation through a combination of goals and beliefs. Then, we design a decision making model based on the social goals and the situational goals of the agent.

Keywords

Virtual agents; social relations; social attitudes; decision making

1. INTRODUCTION

Nowadays machines own a prominent place in our life. Studies show that we are becoming more comfortable with the idea of interacting with them [35] in particular when the interaction involves virtual humans as partners. Our purpose is to improve the naturalness and the believability of these human-machine interactions by integrating socio-emotional components into virtual agents. These components should allow agents to manage the inherently social and emotional dimension of the interaction by expressing an emotional behavior and a social attitude relevant to the context of the interaction.

In [33], Scherer defines social attitudes (also called inter-personal stance) as an *"affective style that spontaneously develops or is strategically employed in the interaction with a person or a group of persons, coloring the interpersonal exchange in that situation"*. Strategical processes can be understood under the light of social roles. Indeed, in particular social contexts or situations, people endow specific roles [10] and intend to express particular behavior and social attitude. Thus, in a restaurant (social context), waiters (social role) are supposed to be nice and polite (social attitude). The same goes for a teacher giving a lesson to a class: they want to be perceived as dominant and controlling the situation.

Spontaneous development of the social attitude can be understood under the light of social relations. Indeed, the social relation between two people influences their behaviour and their social attitudes. For instance, two people who like each other spontaneously tend to comply with the other's requests thus showing a friendly attitude [8]. Moreover, depending on their actions during the interactions, their social relation might change. Hence, their social attitudes have to follow this evolution: if the same people end up hating each other, they will probably express more hostility than they did before.

In order to give the capability to a virtual agent to reason about its role and its own social relation toward the user, and to express an adequate social attitude according to the situation, we designed a model of social decision making. First, we propose a formal model of social relations dynamics based on the agent's goals and beliefs. This representation allows the agent to reason about the situation in order to take a decision and to build a strategy influenced by the agent's socio-emotional state. We also propose a model of decision making influenced by the agent's social relations toward its interaction partners and the agent's role in the interaction. The paper is organized as follows. In Section 2, we present the existing models of social relations while our general approach is described in section 3. In Section 4, our formalization of social relation dynamics is introduced. Our model of decision making is explained in Section 5 and illustrated by variations of a teacher/student scenario in Section 6. We conclude and discuss future work in Section 7.

2. RELATED WORKS OF SOCIAL RELATIONS MODELS

In the first section of this related work, we present works modeling the influence of actions and events on an agent's social relations. In the second section, we present works

modeling the influence of social roles on the agent's attitude and behaviour.

Agents with dynamical social relations: understanding the spontaneous development of social attitude

Agents' social relations can be represented by different dimensions. Many existing models use two dimensions defined by the axis of dominance and the axis of liking. However, it is important to note that the definitions of these dimensions may vary across these models. In [20], liking is defined as *"a general and enduring positive or negative feeling about some person"* [20]. This notion is asymmetric [25, 27]. However, in [2] as in [15], liking involves not only feeling but also commitment. There are also different interpretations of dominance. In [2] and [27] dominance is defined as a hierarchical framework while [25] defines it as "*the capacity of one agent to affect the behavior of another*". This influence is itself characterized by the resources and strategies available to the agent [29].

One approach to model the dynamics of agent's social relations is based on the emotions felt by the agent. In SCREAM [28] the emotions felt by the agent play an important role, changing the relationship according to their valence and intensity. A positive emotion elicited by another agent will raise the liking value towards it, while a negative emotion will have the opposite effect. The authors also add the notion of familiarity, still a function of the emotions but evolving monotonically: only positive emotions are taken into account. Similar dynamics can be found in [21], where authors describe the influence of particular emotions on liking, dominance and solidarity. For instance, an agent A feeling an emotion of pride elicited by another agent B will improve A's values of dominance and liking toward B. These values are initially defined by the role of the agent. In EVA [16], the relation between the agent and the user is represented by two values of friendliness and dominance. Like [28, 21], these values evolve according to four emotions felt by the agent: gratitude, anger, joy and distress.

In SGD [26], the authors try to team up humans with a group of synthetic characters, using a formal representation of liking and dominance. However, the evolution of these two dimensions does not rely on emotions, but on the content of the interactions between the agents. Socio-emotional actions, such as encouraging or disagreeing with one agent, will have an impact (respectively positive and negative) on its liking value. Instrumental actions, such as enhancing an agent's competence or obstructing one of its problems, will have an impact on its dominance.

Avatar Arena focuses on the appreciation in a scenario in which a user must negotiate a schedule appointment with several agents [30]. Before each session, the appreciation level between agents is fixed (either low, medium, or high), as well as their assumptions about other agents' preferences. According to the Congruity Theory described by Osgood and Tannenbaum [23], when an agent discovers a mismatch between its assumption about another agent's preference and what this agent's preference actually is, the former's appreciation toward the latter's decreases. Some other works rely on stage models to implement the notion of intimacy in their agents. This is the case for Laura who encourages users to exercise on a daily basis [5]. Laura's behavior is driven by its intimacy level with users that evolves during the interactions. The more the user interacts with Laura, the more familiar it will behave.

In most of these works, the agent directly expresses the social relation it feels. There is no strategical process, hence no social context adaptation. Moreover, the social relation rarely influences the decision taken by the agent.

Agents with social roles: modeling the strategical processes involved in social attitude

To enhance the believability of virtual agents, several models have proposed to define the agent's role in a particular social context along different variables. These works focus on the influence of these social variables on the agent's behaviour. In [36], the linguistic style and the politeness of a virtual character is determined according to its social distance and power. In Rousseau and Hayes-Roth [31] the agent's behavior is computed based on its personality and attitudes. Gratch [11] proposes a social layer that manages communication and biases plan generation and execution in accordance with a social context . However in these different works the social variables do not vary along the interaction. They are fixed for each agent.

Callejas et al. rely on a circumplex representation (liking, dominance) of social roles to build a computational model of social attitudes for a virtual recruiter in a job interview scenario [6]. In this work, the social attitude of the recruiter is dynamically computed according to the difficulty level of the interview and the anxiety level of the user. The recruiter will be friendly in lower difficulty levels, but might change its attitude as the difficulty increases. Here, the attitude is expressed strategically, in order to comfort or challenge the user. A virtual recruiter is also modeled in [1]. However, in this work, the social attitude expressed by the recruiter depends on its emotions, its mood, and the social cues that the agent gets from the user.

Another example of relational agent is Rea who adapts its dialog strategy according to the principle of trust [4]. Endowing the role of a real estate agent, Rea uses small-talks to enhance the confidence of the user. Once the user becomes more confident with Rea, task-oriented dialog can take place.

Although these works model the influence of the social context on the decision making, most of them rely on static variables: the agent does not adapt its behaviour regarding the other interactants actions. We propose a model of decision making combining dynamic social relations and agent's social role. Thus, the agent plans its strategy taking into account the social context of the interaction, and is able to adapt it depending on the other's actions.

3. GENERAL APPROACH

Our purpose is to build a virtual agent able to reason about its social role and its social relation towards its interactant. The action of the interactant influences the agent's emotional state which in turn influences its social relation toward the interactant. Both emotional state and social relation are part of the agent's beliefs. During an interaction, the agent selects its next action according to its goals and beliefs (including emotional state and social relation). Thus, the action resulting from this reasoning process conveys an adequate social attitude computed from its social role and social relation, itself influenced by the agent's emotional state. Our social decision making model is illustrated by Fig.1 and is described as follows:

Social role and social relation The social role of the agent represents its ideal self in a particular social context.

First, it defines the agent's function during the interaction: the agent might be a teacher, a student, a colleague... It also defines the *ideal state* of the relationship for the agent toward its interactant given the situation: Does the agent want to be an authoritative teacher toward its student or a compliant one? Does the agent want to play the role of an hostile student toward its teacher or a friendly one? As in [6, 1], we define the *ideal social relation* of an agent i toward another agent j in a particular social context using two dimensions of liking and dominance.

$$IdealRelation_{i,j} = < IdealLik_{i,j}; IdealDom_{i,j} >$$
$$\text{with } IdealLik_{i,j} \in [-1; 1] \text{ and } IdealDom_{i,j} \in [-1; 1] \quad (1)$$

These two values are static as the agent maintain its role during the interaction and thus, its ideal relation.

Similar to the *ideal social relation*, we define the *actual social relation* of an agent i toward another agent j using the two dimensions of liking and dominance.

$$ActualRelation_{i,j} = < Lik_{i,j}; Dom_{i,j} >$$
$$\text{with } Lik_{i,j} \in [-1; 1] \text{ and } Dom_{i,j} \in [-1; 1] \quad (2)$$

Unlike the ideal social relation, values of dominance and liking related to the agent's actual social relation i toward the agent j are dynamic. They represent the *actual state* of the relationship. Indeed, i's social relation evolves during the interaction depending on i's emotions (see sect.4). Our social decision making model relies on three kinds of goals: (1) one long-term relational goal, (2) social goals and (3) situational goals.

$$Goals_{i,j} = \{LgtG_{i,j}; SocG_{i,j}; SitG_i\} \quad (3)$$

Long-term relational goal The long-term relational goal describes the importance granted by an agent to maintain a long-term relation with its interactant. This goal is based on the interpersonal complementarity theory [32]. This theory is defined using the circumplex representation of Argyle's dimensions and states that two people with complementary behaviors will experience a better and longer relationship. For instance, two agents i and j are complementary if both are friendly, and either i or j is dominant. On the opposite, if i is friendly and j is hostile, or if both i and j are dominant, the relationship is less likely to work. Moreover, the importance $k \in [0; 1]$ accorded by agent i to the long-term relational goal depends on the liking value of i toward j: the more an agent i likes an agent j, the more agent i values maintaining a long-term relationship with j [8].

$$LgtG_{i,j} = G_{i,j}^k((Lik_{i,j} = Lik_{j,i}) \wedge (Dom_{i,j} = -Dom_{j,i}))$$
$$\text{with } k = max(Lik_{i,j}; 0) \quad (4)$$

Social Goals The social goals are related to Theory of Mind (ToM) [3] and represent how agent i would like agent j to feel toward it. For instance, i might want j to like it, or i might want j to consider it as more dominant. These goals are influenced by the ideal and the social relations of agent i:

$$SocG_{i,j} = \{G_{i,j}^{kl}(Lik_{j,i} > 0); G_{i,j}^{kd}(Dom_{j,i} > 0);$$
$$G_{i,j}^{kd'}(Dom_{j,i} < 0)\} \quad (5)$$

We do not consider the goal $G_{i,j}(Lik_{j,i} < 0)$ (agent i desires agent j to dislike it) because if i dislikes j, it will

consider that maintaining the relationship is not important at all. Thus, i will not have the goal to be disliked by j. These social goals are influenced by the agent ideal relation and its actual relation.

(1) Ideal relation: How the agent sees its own role directly influences its social goal. For instance, as a teacher, agent i might want agent j to like it, and to consider it as more dominant.

(2) actual relation: The agent's actual social relation also influences its social goals. Indeed, according to the complementarity theory, agent i might want agent j to have a complementary relation with it. That means that agent i wants agent j to have the same liking relation toward it, but an opposite relation of dominance.

Importances kl, kd and kd' of these goals are computed as follows:

$$kl = max\left(\frac{IdealLik_{i,j} + Lik_{i,j}}{2}; 0\right)$$
$$kd = max\left(\frac{IdealDom_{i,j} + Dom_{i,j}}{2}; 0\right) \quad (6)$$
$$kd' = max\left(-\frac{IdealDom_{i,j} + Dom_{i,j}}{2}; 0\right)$$

Situational Goals Finally, situational goals are defined by how an agent considers its own social role. For instance, if agent i is a teacher, its situational goal might be to make the student j learn and finish its homework. Those goals do not necessarily involve another agent and their importance k depends on how the agent sees its role. Some agent might consider that working is more important than having fun, while others do not share this opinion.

$$SitG_i = \{G_i^k(1); G_i^{k'}(2); G_i^{k''}(3)...G_i^{kn}(n)\} \quad (7)$$

During the interaction, the social relation between i and j might change according to the emotions felt by i. If the relation evolves in a positive way (i likes j more), then i will care more about the relationship. Therefore, i will care more about j's relation toward it. According to the balance between situational and social goals, the agent makes its decision and selects an appropriate action.
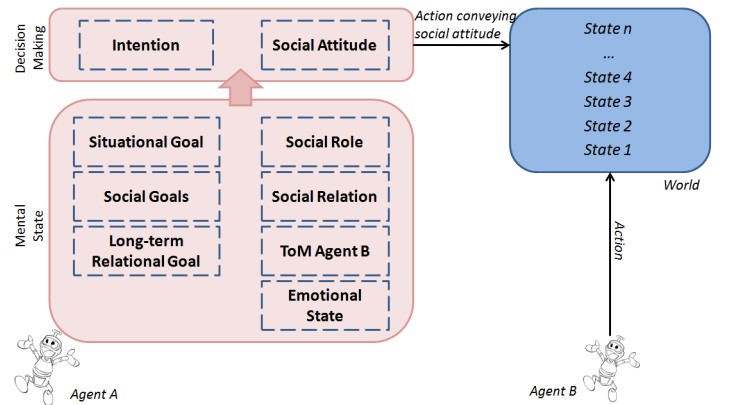

Figure 1: **Global model description**

4. MODEL OF SOCIAL RELATIONS DYNAMICS

We now present our formal model of social relation dynamics. As already mentioned, our representation of the mental state of the agent is defined in terms of goals and beliefs allowing us to model the emotional state of the agent and how this latter affects the social relation of the agent.

4.1 Theoretical Background

Liking dynamics

As explained in sect.2, one approach to model an agent's liking dynamics is based on the emotions it feels. Shiota et al. [34] make a difference between the emotions experienced and those expressed by the agent. The authors describe how distinct positive emotions influence the social relationship depending on its type: parental, romantic, friendship and group. Ortony assesses that emotions felt by an agent play an important role in the evolution of its liking toward others [22]. The direction of the evolution relies on the valence of the emotions felt by the agent: positive emotions increase the degree of liking while negative emotions decrease the degree of liking.

Heider proposes another theory to model liking dynamics [13]. This theory can be represented as a triangular schema between a person P as the focus of the analysis, another person O and an impersonal entity X, which could be a physical object, an idea or an event. If both P and O share the same opinion about X, they like each other more. If P and O have different opinions about X, they like each other less. In their theory, Osgood and Tannenbaum keep the same triangular schema [23]. However, the evolution of the liking between P and O depends on the congruity between P's assumption of O's attitude toward X and O's actual attitude.

Dominance dynamics

Most of the works modeling the dominance of an agent and its dynamics are based on the agent's goals. Emerson [9] assesses that an agent A's level of dependence toward an agent B is *"directly proportional to A's motivational investment in goals mediated by B and inversely proportional to the availability of those goals to A outside of the A-B relation"*. A's level of dominance is then the difference between the two levels of dependence. Castelfranchi [7] formalizes the different patterns of dependence that can happen in a relationship. An agent is dependent on another one if the latter can help the former to achieve one of its goals. The dependence level may vary if the dependent agent finds alternative solutions, or manages to induce a mutual or reciprocal dependence. Hexmoor et al. [14] address autonomy, power and dependence from another perspective. In their work, the agent's power is characterized as a difference between personal weights and *'liberties'* of preferences. The weights influence the agent towards individual or social behavior. The *'liberties'* represent internal or external processes that influence the agent's preferences of choice.

4.2 The logical framework

As mentioned before (see Fig. 1), the agent's mental state is made of its goals and beliefs (including its social relation and emotional state). To formalize the agent's mental state, we introduce the following logical framework:

Let $\neg$, $\wedge$, and $\vee$ represent respectively logical negation, conjunction and disjunction, $\exists$ and $\forall$ existential and universal quantifier. Let the symbols i and j represent two different agents (virtual or real), a and a' different actions, ϕ and ψ logical formulas. As already stated in section 3, $Lik_{i,j}$ and $Dom_{i,j}$ represent respectively the value of liking and dominance of agent i toward agent j. The agent's beliefs and goals are formalized respectively by the modal operator B and G.

- $B_i^h(\phi)$ can be read as the agent i believes that ϕ is true with a strength equal to $h \in]0,1]$.

- $G_i^k(\phi)$ means that the agent i has the goal ϕ and wants ϕ to be true. $G_i^k(\phi) \in SitG_i$ is a situational goal and the importance accorded to this goal is equal to $k \in]0,1]$.

We also define the following temporal operators and two syntactic abbreviations:

- $After_a(\phi)$ means that ϕ is true after the action a

- $Before_a(\phi)$ means that ϕ was true before the action a

- $Resp_i(a)$ means that the agent i is the author of the action a

- $Done_i(a, \phi) \equiv Before_a(\neg\phi) \wedge Resp_i(a) \wedge \phi$ means that the agent i did an action a and ϕ was true after that action.

- $CanDo_i(a, \phi) \equiv \neg\phi \wedge \exists a(Resp_i(a) \wedge After_a(\phi))$ means that the agent i can do an action a and ϕ would be true after that.

The operators are based on a Kripke's possible-world semantics with, for each operator, an accessibility relation [17].

Emotions Our work is based on Lorini's formalization of emotions [18] and [12]. The intensity l of the emotions of gratitude and anger depends on the importance k accorded to the goals. On the other hand, the intensity l of the expectation-based emotions hope and fear [18] relies on the importance k given to the goals G and to the strength h of the beliefs B on the other agent's action. Finally the intensity l of the emotions relief, fearconfirmed, disappointment and satisfaction is related to the intensity of the expectation-based emotions.

$$\begin{cases} Gratitude_{i,j}^l(a,\phi) \equiv G_i^k(\phi) \wedge Done_j(a,\phi) \\ Anger_{i,j}^l(a,\phi) \equiv G_i^k(\neg\phi) \wedge Done_j(a,\phi) \\ \text{with } l = k \end{cases} \quad (8)$$

$$\begin{cases} Hope_{i,j}^l(a,\phi) \equiv G_i^k(\phi) \wedge B_i^h(CanDo_j(a,\phi)) \\ Fear_{i,j}^l(a,\phi) \equiv G_i^k(\neg\phi) \wedge B_i^h(CanDo_j(a,\phi)) \\ \text{with } l = \dfrac{k+h}{2} \end{cases} \quad (9)$$

$$\begin{cases} Relief_{i,j}^l(a,\phi) \equiv Fear_{i,j}^l(a,\phi) \wedge Done_j(a,\neg\phi) \\ FearConfirmed_{i,j}^l(a,\phi) \equiv Fear_{i,j}^l(a,\phi) \wedge Done_j(a,\phi) \\ Disappointment_{i,j}^l(a,\phi) \equiv Hope_{i,j}^l(a,\phi) \wedge Done_j(a,\neg\phi) \\ Satisfaction_{i,j}^l(a,\phi) \equiv Hope_{i,j}^l(a,\phi) \wedge Done_j(a,\phi) \end{cases} \quad (10)$$

4.3 Liking and Dominance dynamics

Liking Dynamics Our model of liking dynamics is based on [22]. As explained in section 4.1, a positively valenced emotion of an agent i triggered by an agent j has a positive influence on i's liking toward j. On the opposite, a negative emotion of i triggered by j lowers i's liking toward j. The liking dynamics not only depends on the valence of the emotion felt by i, but also depends on the intensity l of this emotion. The higher the intensity l of the emotion, the more it influences i's liking toward j. The liking dynamics is captured by the following formulas:

$$\left.\begin{array}{r} Gratitude_{i,j}^{l}(a,\phi) \\ Hope_{i,j}^{l}(a,\phi)) \\ Satisfaction_{i,j}^{l}(a,\phi)) \\ Relief_{i,j}^{l}(a,\phi) \end{array}\right\} Lik_{i,j} = min(Lik_{i,j} + l; 1) \quad (11)$$

$$\left.\begin{array}{r} Anger_{i,j}^{l}(a,\phi) \\ Fear_{i,j}^{l}(a,\phi) \\ FearConfirmed_{i,j}^{l}(a,\phi) \\ Disappointment_{i,j}^{l}(a,\phi) \end{array}\right\} Lik_{i,j} = max(Lik_{i,j} - l; -1) \ (12)$$

According to the equations 8,9 and 10 introduced in section 4.2, the evolution of i's social relation toward j can be represented by a combination of goals and beliefs. For example, if j helps i to achieve one of i's goals, i feels gratitude towards j and its liking toward j increases.

Dominance dynamics Our model of dominance dynamics is rooted in [9]. An agent i is less dominant toward an agent j if i believes that j can influence one of i's goals. The dominance dynamics also depends on the importance k accorded by i to its goal, and the strength of its belief about j's action. The more i considers its goal as important, the less dominant i is. Moreover, the more i believes that j can influence its goal, the less dominant i is. According to the equations 8,9 and 10, we obtain the following formulas:

$$\left.\begin{array}{r} Fear_{i,j}^{l}(a,\phi) \\ Hope_{i,j}^{l}(a,\phi) \\ Satisfaction_{i,j}^{l}(a,\phi) \\ FearConfirmed_{i,j}^{l}(a,\phi) \end{array}\right\} Dom_{i,j} = max(Dom_{i,j} - l; -1) \ (13)$$

$$\left.\begin{array}{r} Relief_{i,j}^{l}(a,\phi) \\ Disappointment_{i,j}^{l}(a,\phi) \end{array}\right\} Dom_{i,j} = min(Dom_{i,j} + l; 1) \ (14)$$

For instance, agent i **dominance** toward j **decreases** if i fears that j can prevent one of i's goals.

5. MODEL OF SOCIAL DECISION MAKING

5.1 PsychSim implementation

We implemented our model in PsychSim [19], where agents are decision-theoretic models with actions, goals and beliefs. Moreover, PsychSim agents have the notion of mental horizon whereby an agent can project into the future, simulating the consequences of their actions. As part of this projection, an agent can take into account how other agents will react, including how those agents' beliefs will change. To perform this projection, PsychSim agent incorporates Theory of Mind. Specifically an agent can have mental models of other agents that are fully specified agent models with beliefs, goals and actions, thus allowing an agent to simulate other agents. Through the dynamic changes of the agent's mental state, the emotional state of the agent is implicitly captured. Thus the evolution of the social relation of the agent is influenced by its emotional state.

Representing actual social relation and ideal social relation as states of the world. States are objective facts about agent's status or about the world itself. The features of a state can be booleans, finite sets of possible values, or a range of continuous values. In PsychSim, an agent only has indirect knowledge about the world, therefore partial observability. Thus an agent's subjective beliefs about the world may differ from the actual world. As explained in section 3, we represent the agent's *actual social relation* and the *ideal social relation* associated with its role using two dimensions of dominance and liking. For each agent, ideal liking, ideal dominance, actual liking and actual dominance are represented as state features ranging from -1 to 1.

Actions. In PsychSim, agent's actions constitute the possible decisions an agent can make. In particular, actions can be either physical actions or dialog acts whose influence is largely on other agents' beliefs. Each action is defined by an actor and a name but might also include a target and a parameter that modifies the action impact. Actions dynamics represent the influence of actions on the states variables introduced above. For instance, shouting at the student might have a negative influence on the student's liking value. Traditional notions of action preconditions and effects are realized through these dynamics. We saw in section 4 how agent j's actions influence agent i's social relation through the emotions i feels. We do not explicitly model agent i emotional state in PsychSim. However, according to the effects of j's actions on agent i's goals, i's social relation evolves. For instance, j's action helping i to achieve its goal enhances i's liking toward j. The evolution of the liking feature associated with one action relies on the following equation:

$$Lik_{t+1} = Lik_t + (LikMaxValue - Lik_t) * weight \quad (15)$$

Indeed, repeating the same action decreases the effect of this action on the other's liking. The weight represents the importance accorded by i to the goal impacted by the action. As stated in section 4, influencing important goals brings emotions with high intensity, thus having a high influence on social relation.

Representing goals as reward functions. The reward function maps the state of the world into an evaluation of the benefit to the agent. In this work, we use a simplified version whereby the agent has relative preferences for goals defined as the agent's desire to maximize or minimize particular state variables. For instance, agent i has the goal to maximize its fun. Moreover, the agent might also want to alter other agents' states or beliefs. For instance, the agent i wants to maximize j's knowledge. Each goal is associated with a preference weight representing the importance accorded by the agent to this goal. According to this weight, each action will be attributed an expected utility: actions having a positive (respectively, negative) influence on a high

importance goal are considered more useful (respectively, useless). The situational goals of the agent are then represented by different reward functions related to states of the world. The importance accorded to these situational reward functions is fixed. On the opposite, computation of social goals importance relies on the four social states described above following the formula 6.
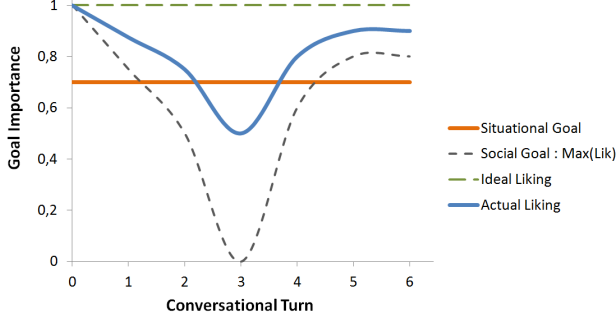

Figure 2: The balance between situational goals and social goals relies on the social relations dynamics. Situational goals are fixed while social goals vary. In this example, for a maximal ideal liking, social goal (here maximizing the liking of the other toward the agent) becomes less important than the situational goal as the liking value drops.

Representing beliefs strength as mental models distribution. An agent in PsychSim has not only beliefs about itself, but also about the other agents. When planning its own actions, the agent takes all these beliefs into consideration. Each agent has a mental model for itself, and at least one mental model for each other agent (mental models include agent's state, beliefs, goals, and actions). For instance, a tutor might have two different models for a student: one where the student is motivated by finishing its homework (model A), and another one where the student just wants to have fun (model B). The agent's belief about another agent is represented as a probability distribution of these mental models. For instance, the tutor might initially be sure at 75% that the student wants to work (model A), and 25% that the student wants to play (model B). In PsychSim, the agent is able to infer another agent's goals according to the latter's actions. Hence, during the interaction, the tutor might notice that the student's behavior does not fit this distribution (for instance, if the student keeps playing although it is asked to do its homework). Therefore, the tutor will update its belief, and will change the distribution according to this observation. When computing the dominance value of an agent i, this probablity is taken into account: the more an agent i believes that agent j can do an action that will influence its goals, the less dominant i feels toward j. The importance accorded to the goal also influences the final value of dominance

$$Dom_{t+1} = Dom_t + (DomMaxValue - Dom_t)$$
$$* \frac{weight + belStrength}{2}$$

Repeating the same action decreases the effect of this action on the other's dominance. This is coherent with Raven's theory [29], assessing that multiple use of reward/coercive strategies lowers the impact of these strategies on others.

Horizon and decision making. As chess players, agents are able to project themselves few steps forward and thus predict the outcome of all the possible sequences of actions. Indeed, when an agent plans its next action, it first projects itself into the future to evaluate the effect of each of its possible actions on the different states. Then, using its mental models of the other agents, it will predict their expected actions and their impact on their states. Then again, the agent will anticipate its reaction, and so on. The number of steps that the agent is able to predict is called *horizon*. Hence, an agent with a high horizon is able to look several steps in the future, and to predict outcomes that an agent with a low horizon is not able to foresee. When the agent finishes its projection, it evaluates the overall effect of each sequence of actions according to its goals. Then, the agent selects the action with the highest expected utility.

6. USE CASE: TUTOR-STUDENT INTERACTION

To illustrate our model, we define a scenario taking place in the following social context: one agent *Bob* endowing the role of a virtual tutor wants to make another agent *Alice* do her homework. As shown in [37], two teachers expressing different social attitudes might get different results when giving a class. Moreover, recent studies have shown that a good tutor/student relationship may improve the student's motivation and her academic achievements [24]. Hence, modeling an agent trying to make a student work presents an interesting challenge.

Tutor's goals and actions Being a tutor, the agent *Bob* has only one situational goal $SitG_{Bob}$, namely the student finishes her homework (operationalized as working for two units of time) during the interaction. Agent *Bob* can do several actions (including doing nothing) to reach his goal.

$$SitG_{Bob} = G_{Bob}(HomeworkDone)$$

$RequestWork$ means that *Bob* informs *Alice* that he wants *Alice* to work and to do her homework.

$$Request_{Bob}(Alice, HomeworkDone)$$
$$\Rightarrow B_{Alice}(G_{Bob}(Done_{Alice}(Work, HomeworkDone)))$$

$PromiseToPlay$ means that *Bob* requests *Alice* to finish her homework and that he would play with her if she satisfies that request.

$$Promise_{Bob}(Alice, HomeworkDone, PlayWithStudent)$$
$$\Rightarrow Request_{Bob}(Alice, HomeworkDone) \wedge$$
$$Inform_{Bob}(Alice, CanDo(Bob, PlayWithStudent))$$

$ThreatenToSwitchOffConsole$ means that *Bob* requests *Alice* to work and that he would switch off the console if she does not.

$$Threaten_{Bob}(Alice, HomeworkDone, SwitchOffConsole)$$
$$\Rightarrow Request_{Bob}(Alice, HomeworkDone) \wedge$$
$$Inform_{Bob}(Alice, CanDo(Bob, SwitchOffConsole))$$

$PlayWithStudent$ means that *Bob* can increase *Alice*'s fun by playing with her.

$$PlayWithStudent_{Bob}(Alice)$$
$$\Rightarrow Done_{Bob}(PlayWithStudent, Fun)$$

$SwitchOffConsole$ means that Bob can switch off the console. $Alice$ will not be able to play anymore with the console after that.

$$SwitchOffConsole_{Bob}(Alice)$$
$$\Rightarrow Done_{Bob}(SwitchOffConsole, ConsoleSwitchedOff)$$

Student's goals and actions Being a student, $Alice$ has two different situational goals $SitG_{Alice}$: she wants to play but also wants to do her homework. She can do three different actions (including doing nothing).

$$SitG_{Alice} = \{G_{Alice}(HomeworkDone), G_{Alice}(Fun)\}$$

$Play$ means that $Alice$ increases her level of fun if she plays. This effect decreases as $Alice$ plays repeatedly (it approaches an asymptotic, see equation15).

$$Play_{Alice}(Bob)$$
$$\Rightarrow Done_{Alice}(Play, Fun)$$
with $Fun_{t+1} = Fun_t + (FunMaxValue - Fun_t) * weight$

$Work$ means that $Alice$ works and finishes one exercise.

$$Work_{Alice}(Bob)$$
$$\Rightarrow Done_{Alice}(Work, HomeworkDone)$$

Finally, we built four different models of the student $Alice$ depending on relative preference for her goals (Homework-Done Vs PlayVideoGames) and her horizon (high Vs low). We ran several simulations, varying these different parameters: (1) social goals, and (2) horizon. In the following subsection, we describe some of these simulations and analyze the results we obtained.
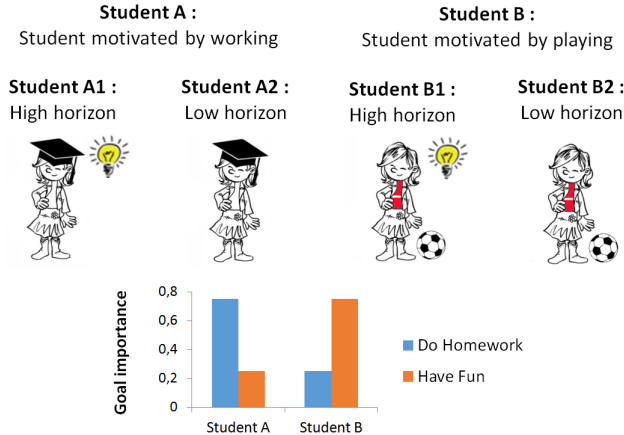

Figure 3: Different models of the student. Student A considers doing homework as more important than having fun, while student B prefers to have fun, regardless their horizon.

6.1 Simulation 1: Reference simulation

In this simulation, we consider the influence of social relation and social role during the interaction and how the interaction might change the balance between situational and social goals. We expect that the tutor will adapt his strategy depending on the student reactions.

Starting situation: Student B2

	Tutor	Student
Ideal Relation	IdealLik(High) IdealDom(Med)	IdealLik(Neutral) IdealDom(Neg)
Actual Relation	Lik(Med) Dom(Med)	Lik(Neutral) Dom(Low)
Situational Goals	Work(HighImp)	Work(LowImp) Fun(HighImp)
Social Goals	MaxLik(HighImp) MinDom(MedImp)	

The teacher Bob wants the student $Alice$ to do her homework, and also wants to maintain the relationship with her. He believes he is dominant toward the student, and his liking toward her is moderately high. The student strongly desires to play, and considers that doing her homework is less important. Her liking is neutral, and she believes she is slightly dominant toward the teacher.

Actions sequence

Step 1: Student-Play

Student $Alice$ believes she is dominant. Indeed, with her low horizon, $Alice$ does not believe that the tutor Bob can influence her goals. Therefore, $Alice$ plays because it is more desirable for her.

Step 2: Teacher-PromiseToPlay

The tutor Bob wants to maintain his relationship with the student $Alice$. He then considers a reward strategy, and promises $Alice$ that he will play with her if she finishes her homework first. This way, it should lower $Alice$'s dominance while enhancing her liking, thus making her work.

Step 3: Student-Play

$Alice$ still believes she is dominant. She does not believe she will be rewarded if she works, so she keeps playing.

Step 4: Teacher-PromiseToPlay

Bob still considers that maintaining a good relationship with $Alice$ is more important than forcing her to work. Bob tries again to convince his student using a reward strategy.

Step 5: Student-Play

$Alice$ still believes she is dominant. she does not believe she will be rewarded so she keeps playing.

Step 6: Teacher-SwitchOffConsole

Since $Alice$ keeps playing, the tutor gets disappointed again (Bob expected $Alice$ would do her homework, but she did not fulfill Bob's expectations). Bob likes $Alice$ less (decrease of his liking), and considers now that maintaining the relationship with her is less important than finishing homework. Thus, the tutor changes his strategy, and uses his coercive power to switch off $Alice$'s console.

Step 7: Student-Work

$Alice$ now believes the tutor is dominant. She still believes that playing is more important for her than working, but she works because she has no other options.

Step 8: Teacher-DoNothing

Bob believes that $Alice$ will keep working now that her console is switched off. Bob has nothing else to do.

Step 9: Student-Work

$Alice$ still has nothing to do but work.

Ending situation

After the interaction, we notice that the relationship between the tutor and his student is not so bad. The tu-

	Tutor	Student
Ideal Relation	IdealLik(High) IdealDom(Med)	IdealLik(Neutral) IdealDom(Neg)
Actual Relation	Lik(Med) Dom(Med)	Lik(Low) Dom(Neg)
Situational Goals	Work(HighImp)	Work(LowImp) Fun(HighImp)
Social Goals	MaxLik(HighImp) MinDom(MedImp)	

tor *Bob* eventually managed to make *Alice* work, so his liking toward her slightly increases and he still feels dominant. Since *Bob* tried to use reward strategies first, *Alice* also slightly increases her liking toward him: the reward strategy increased liking, offsetting the following decrease in liking due to switching off the console. However, *Alice* does not feel dominant anymore. Her dominance level has dropped.

6.2 Simulation 2: Playing with goals

In this simulation, the starting situation is the same as the reference one (see section 6.1) except that we adjusted the ideal social relation of the tutor *Bob* so that he does not have any social goal ($SocLik_{Bob,Alice} + Lik_{Bob,Alice} \leq 0$). By doing this, *Bob* only focuses on his situational goal: he will choose the most efficient plan that will make *Alice* finish her homework.

Actions Sequence
Step 1: Student-Play.
Step 2: Teacher-ThreatenToSwitchOffConsole.
Step 3: Student-Play.
Step 4: Teacher-SwitchOffConsole.
Step 5: Student-Work.
Step 6: Teacher-DoNothing.
Step 7: Student-Work.
Step 8: Teacher-DoNothing.
Step 9: Student-Work.

Student *Alice* believes she is dominant. Indeed, with her low horizon, *Alice* does not believe that the tutor *Bob* can influence her goals. Therefore, *Alice* plays because it is more desirable for her. The tutor does not have the goal to maintain the relationship, so he chooses a coercive strategy. Indeed, by switching off *Alice*'s console, thus forbidding her to play, *Bob* knows that she will have nothing to do but work.

After the interaction, the relation between *Bob* and *Alice* is bad. The tutor eventually managed to make *Alice* work, so his liking toward her slightly increases and he still feels dominant. However, *Bob* used coercive strategies, threatening *Alice* to switch off her console if she would not do her homework. Even though it was effective, this strategy had a disastrous influence on *Alice*'s relation since *Alice*'s liking value toward *Bob* has noticeably dropped and reached negative value.

6.3 Simulation 3: Playing with horizon

In this simulation, we want to check the effect of the horizon on the scenario. Starting from the same situation as the one described in section 6.1, we only raise the student's horizon. By doing this, we expect different sequences of actions, and different social relations in the end: the student with a

high horizon should be able to delay her gratification. She should be able to resist the temptation of getting immediate fun by playing, and to wait for the late reward of playing with the tutor. (note we could also get a similar effect by giving to the student different discount factors on future rewards).

Actions sequence
Step 1: Student-Work.
Step 2: Teacher-DoNothing.
Step 3: Student-Work.
Step 4: Teacher-DoNothing.
Step 5: Student-Work.
Step 6: Teacher-PlayWithStudent.
Step 7: Student-Play.
Step 8: Teacher-PlayWithStudent.
Step 9: Student-Play.

Since she has a high horizon, the student *Alice* is able to project herself more steps ahead into the future. She believes that the teacher *Bob* will switch off her console if she does not work. *Alice* also believes that *Bob* will play with her if she finishes her homework (which she expects will take two units of working time). After considering all the possibilities, *Alice* decides that working is worth it. Indeed, it will bring her more fun to play with *Bob* rather than not working and playing alone. Moreover, *Alice* will also complete her other goal of finishing her homework. Therefore, she decides to delay her gratification (playing gives *Alice* a better immediate reward than working) and start working. *Bob* does not have to do anything at the beginning because *Alice* is working. Once *Alice* finishes her homework, *Bob* starts to play with her and keeps playing with her until the end of the interaction.

In the end, the relationship between *Bob* and *Alice* is good. *Alice* fulfilled *Bob*'s goal by completing her homework, and *Bob* did not do anything but play with her, improving her fun. Thus, the liking value of both agents is very high. The dominance value did not change during the interaction.

7. CONCLUSION AND FUTURE WORK

In this paper, we presented an agent able to express a particular social attitude that takes into account spontaneous and strategical processes. Our main contributions are: (1) a formal model of dynamical social relations relying on goals and beliefs, and (2) a social decision making model based on the agent's social relation and social role. We have illustrated our model on a tutor-student use-case. We also noticed that if at least one agent had a low horizon, the relationship would decrease. Indeed, agents with low horizon are not able to consider the affective consequences of their actions, on others as well as on themselves. By analogy to humans, agents lacking Socio-Emotional Intelligence can not maintain a proper relationship. Our next step is to evaluate our model. After generating different tutor/student scenarios, we plan to ask participants whether they recognize the teacher's social attitude through its sequence of actions. Our work still has some limits. For instance, we did not address the following problem: should *Bob* like *Alice* more if he believes that *Alice* did her homework because she was forced to do so (in terms of attribution theory, she is not praiseworthy if she was forced).

8. ACKNOWLEDGEMENT

REFERENCES

[1] K. Anderson, E. André, T. Baur, S. Bernardini, M. Chollet, E. Chryssafidou, I. Damian, C. Ennis, A. Egges, P. Gebhard, et al. The tardis framework: intelligent virtual agents for social coaching in job interviews. In *Advances in Computer Entertainment*, pages 476–491. Springer, 2013.

[2] D. Ballin, M. Gillies, and I. Crabtree. A framework for interpersonal attitude and non-verbal communication in improvisational visual media production. pages 203–210, London, UK, 2004.

[3] S. Baron-Cohen, A. M. Leslie, and U. Frith. Does the autistic child have a 'theory of mind'? *Cognition*, 21(1):37–46, 1985.

[4] T. Bickmore and J. Cassell. Social dialogue with embodied conversational agents. In *Advances in natural multimodal dialogue systems*, pages 23–54. Springer, 2005.

[5] T. W. Bickmore and R. W. Picard. Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 12(2):293–327, 2005.

[6] Z. Callejas, B. Ravenet, M. Ochs, and C. Pelachaud. A computational model of social attitudes for a virtual recruiter. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*, pages 93–100. International Foundation for Autonomous Agents and Multiagent Systems, 2014.

[7] C. Castelfranchi, M. Miceli, and A. Cesta. Dependence relations among autonomous agents.

[8] R. B. Cialdini and N. J. Goldstein. Social influence: Compliance and conformity. *Annu. Rev. Psychol.*, 55:591–621, 2004.

[9] R. Emerson. Power-dependence relations. *American Sociological Review*, 27(1):31–41, 1962.

[10] E. Goffman. Presentation of self in everyday life. *American Journal of Sociology*, 55:6–7, 1949.

[11] J. Gratch. Socially situated planning. In *Socially Intelligent Agents*, pages 181–188. Springer, 2002.

[12] N. Guiraud, D. Longin, E. Lorini, S. Pesty, and J. Rivière. The face of emotions: a logical formalization of expressive speech acts. In *The 10th International Conference on Autonomous Agents and Multiagent Systems-Volume 3*, pages 1031–1038. International Foundation for Autonomous Agents and Multiagent Systems, 2011.

[13] F. Heider. *The Psychology of Interpersonal Relations*. Psychology Press, 1958.

[14] H. Hexmoor. A model of absolute autonomy and power: toward group effects. *Connection Science*, 14(4):323–333, 2002.

[15] K. Isbister. *Better Game Characters by Design, A Psychological Approach*. Morgan Kaufmann, 2006.

[16] Z. Kasap, M. Ben Moussa, P. Chaudhuri, and N. Magnenat-Thalmann. Making them remember, emotional virtual characters with memory. *Computer Graphics and Applications, IEEE*, 29(2):20–29, 2009.

[17] S. Kripke. Semantic considerations on modal logic. *Acta Philosophica Fennica*, 16:83–94, 1963.

[18] E. Lorini. A dynamic logic of knowledge, graded beliefs and graded goals and its application to emotion modelling. In *Logic, Rationality, and Interaction*, pages 165–178. Springer, 2011.

[19] S. C. Marsella, D. V. Pynadath, and S. J. Read. Psychsim: Agent-based modeling of social interactions and influence. In *Proceedings of the international conference on cognitive modeling*, pages 243–248. Citeseer, 2004.

[20] L. Moshkina and R. Arkin. On tameing robots. *IEEE International Conference on Systems, Man and Cybernetics*, 4:3949–3959, 2003.

[21] M. Ochs, N. Sabouret, and V. Corruble. Simulation of the dynamics of nonplayer characters' emotions and social relations in games. *Computational Intelligence and AI in Games, IEEE Transactions on*, 1(4):281–297, 2009.

[22] A. Ortony. Value and emotion. *Memories, thoughts, and emotions: Essays in the honor of George Mandler*, pages 337–353, 1991.

[23] C. E. Osgood and P. H. Tannenbaum. The principle of congruity in the prediction of attitude change. *Psychological review*, 62(1):42, 1955.

[24] H. J. Pennings, J. van Tartwijk, T. Wubbels, L. C. Claessens, A. C. van der Want, and M. Brekelmans. Real-time teacher–student interactions: A dynamic systems approach. *Teaching and Teacher Education*, 37:183–193, 2014.

[25] R. Prada and A. Paiva. *New Advances in Virtual Humans*, volume 140, chapter Social Intelligence in Virtual Groups, pages 113–132. Springer, 2008.

[26] R. Prada and A. Paiva. Teaming up humans with autonomous synthetic characters. *Artificial Intelligence*, 173(1):80–103, 2009.

[27] H. Prendinger, S. Descamps, and M. Ishizuka. Scripting affective communication with life-like characters in web-based interaction systems. *Applied Artificial Intelligence: An International Journal*, 16(7-8):519–553, 2002.

[28] H. Prendinger, S. Descamps, and M. Ishizuka. Scripting affective communication with life-like characters in web-based interaction systems. *Applied Artificial Intelligence*, 16(7-8):519–553, 2002.

[29] B. Raven. The bases of power and the power/interaction model of interpersonal influence. *Analyses of Social Issues and Public Policy*, 8(1):1–22, 2008.

[30] T. Rist, M. Schmitt, C. Pelachaud, and M. Bilvi. Towards a simulation of conversations with expressive embodied speakers and listeners. In *Computer Animation and Social Agents, 2003. 16th International Conference on*, pages 5–10. IEEE, 2003.

[31] D. Rousseau and B. Hayes-Roth. A social-psychological model for synthetic actors. In *Proceedings of the second international conference on Autonomous agents*, pages 165–172. ACM, 1998.

[32] P. Sadler and E. Woody. 8 interpersonal complementarity. *Handbook of interpersonal psychology: Theory, research, assessment, and therapeutic interventions*, page 123, 2010.

[33] K. Scherer. What are emotions? and how can they be measured? *Social Science Information*, 2005.

[34] M. N. Shiota, B. Campos, D. Keltner, and M. J. Hertenstein. Positive emotion and the regulation of interpersonal relationships. *The regulation of emotion*, pages 127–155, 2004.

[35] S. Turkle. *Close Engagements with Artificial Companions: Key social, psychological, ethical and design issues*, chapter In Good Company: On the treshold of robotic companions, pages 1–10. John Benjamins, 2010.

[36] M. A. Walker, J. E. Cahn, and S. J. Whittaker. Improvising linguistic style: Social and affective bases for agent personality. In *Proceedings of the first international conference on Autonomous agents*, pages 96–105. ACM, 1997.

[37] T. Wubbels, M.-C. Opdenakker, and P. Den Brok. Let's make things better. In *Interpersonal relationships in education*, pages 225–249. Springer, 2012.