Social Role Awareness in Animated Agents

Helmut Prendinger
Department of Information and Communication Engineering
School of Engineering, University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
+81 3 5841 6755
helmut@miv.t.u-tokyo.jp

Mitsuru Ishizuka
Department of Information and Communication Engineering
School of Engineering, University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
+81 3 5841 6347
ishizuka@miv.t.u-tokyo.jp

ABSTRACT

This paper promotes *social role awareness* as a desirable capability of animated agents, that are by now strong affective reasoners, but otherwise often lack the social competence observed with humans. In particular, humans may easily adjust their behavior depending on their respective role in a socio-organizational setting, whereas their synthetic pendants tend to be driven mostly by attitudes, emotions, and personality. Our main contribution is the incorporation of 'social filter programs' to mental models of animated agents. Those programs may qualify an agent's expression of its emotional state by the social context, thereby enhancing the agent's believability as a conversational partner or virtual teammate. Our implemented system is entirely web-based and demonstrates socially aware animated agents in an environment similar to Hayes-Roth's Cybercafé.

Keywords

believability, social agents, human-like qualities of synthetic agents, social dimension in communication, affective reasoning and emotion expression

1. INTRODUCTION

Ever since Bates and Reilly promoted *believable* agents in their 'Oz project' [2], there has been continued interest to give animated agents the illusion of life. It is now widely accepted that emotion expression and personality are key components of believable agents. Moreover, Cassell and her coworkers [4] recently provided convincing evidence of the importance of non-verbal 'embodied' (conversational) behavior for believable lifelike agents. Animated agents with believable behavior are used as virtual tutors in interactive learning environments (e.g., Johnson *et al.* [16]), as virtual presenters on the web (e.g., André *et al.* [1], Ishizuka *et al.* [14]), and as virtual actors for entertainment (e.g., Rousseau and Hayes-Roth [28]). Although those agents achieve convincing behavior in their respective predefined roles, they might fall short of 'social robustness' when put into a different scenario. E.g., a tutor agent might be believable for a student (in the teacher role) but not in the role of a peer. Humans, on the other hand, are always aware of the roles they play in a certain social setting and typically behave accordingly.

In this paper, we will argue that *social role awareness* is an important feature of human-human communication which should be integrated to existing animated agents approaches. Social role awareness is easily illustrated, as in the following conversation.

Aspirant (to secretary):	*I need a copy of this document.*
Secretary (to aspirant):	*So what?*
Manager (to secretary):	*I need a copy of this document.*
Secretary (to manager):	*Sure. I will do it immediately.*

The secretary's polite behavior towards the manager is consistent with her unfriendly reaction to the aspirant, simply because she is aware of her role-specific rights and duties in the socio-organizational context of an office environment. Awareness of her role as secretary allows her to ignore the (indirect) request from the aspirant (whose role is assumed to be lower than the secretary's on the power scale). We believe that the conversation above cannot be understood (or generated) by means of reasoning about personality and attitudes alone. Even if the secretary happens to have a rather aggressive personality and she does not like her manager, she would respect the manager's rights and obey to his or her request. Similarly, we will argue that an agent's role determines its way of emotion expression. Consider a situation where the (aggressive) secretary is angry with her manager. She will presumably not show her emotion to the manager, being aware of her social role as an employee. On the other hand, she might express her anger to a fellow secretary who has equal social power.

Our goal is to create autonomous agents that can serve as dramatically interesting conversational partners for the task of web-based language conversation training. Specifically, the animated agent approach will be used to improve English conversation skills of native speakers of Japanese. Since interactions are set up as role-playing dramas and games, strong requirements are imposed on the agents' social abilities. Social reasoning will be blended with a rather standard theory of reasoning about emotion (Ortony *et al.* [25, 24]) and a simple model of personality. We employ Moulin and

AGENTS'01, May 28-June 1, 2001, Montréal, Quebec, Canada.

Rousseau's [22] approach to model and simulate conversations, which provides a rich framework for many aspects of inter-agent communication. The programmable interface of the Microsoft Agent package [19] is used to run our role-playing scenarios. Although these off-the-shelf agent characters are quite restricted in the number of behaviors, the package comes ready with a speech recognizer and text-to-speech engine that allow client-side execution in a web browser.

The rest of this paper is organized as follows. The next section discusses related work. In Section 3, we describe a framework for modeling and simulating conversations. In the following section, we introduce a simple affective reasoner and argue that reasoning about emotion and personality is not sufficient to achieve believable emotion expression. Section 5 describes the basic notions underlying social reasoning and introduces social filter programs that function as a filter between affective state and emotion expression. In Section 6, we first explain the web technology used to run the animated agents. After that, we illustrate our approach by example runs of a Cybercafé-style role-playing drama. Section 7 discusses the paper and suggests further extensions and refinements. Finally, we summarize the paper.

2. RELATED WORK

Several research groups have addressed the problem of socially intelligent agents in the framework of multi-agent systems (e.g., Jennings [15], Castelfranchi [5]). On the other hand, there are relatively few researchers who focus on social role related behavior from the perspective of the believability of animated character agents.

A notable exception is Hayes-Roth and her co-workers (Hayes-Roth *et al.* [13], Rousseau and Hayes-Roth [28]). In [13], role-specific behavior is studied in the context of characters (animated agents) that function as actors in a master-servant scenario. Social roles are defined by behaviors that represent a character's status, e.g., high status is effected through a quiet manner and ways of talking that forbid interrupting. A character's believability in a specific role is justified by adopting guidelines from the literature on the art of drama, rather than by exploiting a social-psychological model. Thus, no attention is paid to a character's representation of its role. Rousseau and Hayes-Roth [28], however, propose an elaborate social-psychological model that considers personality, emotions (moods), and attitudes. Rules are defined that allow to select appropriate behavior depending on a character's personality and attitudes, thereby enhancing the character's believability. Interestingly, the social roles discussed in [13] are only of marginal importance in this model. By contrast, we believe that an agent's awareness of its social role is equally important for action selection, and may even overrule the influence of personality.

Walker *et al.* [29] promote linguistic style as a key aspect for believable agents. Linguistic choices are seen as realizations of agents' personality, and subject to social variables such as social distance between agents or the power one agent has over another. The *Linguistic Style Improvisation* (LSI) framework is based on speech acts theory and a theory of linguistic social interaction. As a theory of social conversation, LSI is clearly superior to our approach, which does not provide a formalization of speech acts and uses a simpler algorithm to decide the linguistic style of utterances. However, we will provide a subtler account of the interaction between an agent's emotional state, personality, social role, and emotion expression.

Gratch [10] introduces 'social control programs' on top of a general purpose planning system. In this system, plan generation and execution are biased by the characteristics of the social context. A so-called 'personality GUI' contains the agent's goals, its social status, etiquette, (in)dependence, and attitudes towards other agents. Besides those static features of the agent's mental state, Gratch [10] introduces dynamic features of the social context such as the communicative state, the plan state, and the agent's emotional state. Social rules encode commonsense rules of social interaction, e.g., help a friend or avoid that some other agent interferes with your plan. For the rather rigid organizational setting in which military commander agents operate, Gratch and Hill [11] introduce social concepts similar to ours. The main difference to our work is that we place social control programs at the interface of the module that reasons about emotion and the module that renders the emotional state to actual behavior. By considering the social context, we aim to achieve believable emotion expression rather than generate socially adjusted plans.

Guye-Vuillème and Thalmann recently started to work on an architecture for believable social agents which is based on four sociological concepts: social norms, values, world view, and social role (see abstract [12]).

Finally, the research most relevant to ours is done by Moulin and collaborators [21, 22]. It will be described in the following section. We also continue work of Prendinger and Ishizuka [27] who motivate the role-playing metaphor for interactive learning environments.

3. A CONCEPTUAL FRAMEWORK FOR SIMULATING CONVERSATIONS

A conversation is typically seen as an activity where multiple (locutor-)agents participate and communicate through multiple channels, such as verbal utterances, gestures and facial display. Each agent has its own goals and will try to influence other participants' mental states (e.g., emotions, beliefs, goals). Moulin and Rousseau [22] distinguish three levels of communication:

- At the *communication level* agents perform activities related to communication maintenance and turn-taking.
- At the *conceptual level* agents transfer concepts.
- At the *social level* agents manage and respect the social relationships that hold between agents.

Our system integrates the second and third level. The communicative level basically implements conversational features of human-human communication, as proposed by Cassell and Thórisson [4]. At the conceptual level, information is passed from one agent to other agents as a (simplified) symbolic representation of the utterance, e.g., if an agent orders a beer, this is simply represented as *order_beer*. According to their role in the social context, the social level puts behavioral constraints on agents' actions and emotion expression (Moulin [21]). This issue will be discussed in detail below.

As an example, consider an agent character playing the role of a customer called 'Al' and an agent character in the role of a waiter called 'James'. Al orders a beer from James by saying “May I order a beer please?”. The corresponding *communicative act* is formalized as

$$com\_act(al,james,order\_beer,polite,happiness,s0)$$

where the argument ‘polite’ is a qualitative evaluation of the linguistic style (LS) of the utterance, the argument ‘happiness’ refers to Al’s emotion expression, and $s0$ denotes the situation in which the utterance takes place.

As in [22], we assume that a conversation is governed by

- a *conversational manager* that maintains a model of the conversation, and
- an *environmental manager* that simulates the environment in which the agents are embedded.

For simplicity, we assume that the conversational manager operates on a shared knowledge base that is visible to all agents participating in the conversation (except for the user). It stores all concepts transferred during the conversation by updating the knowledge base with

$$com\_act(S,H,C,LS,E,Sit)$$

facts. The resulting ‘model’ of the conversation will eventually be substituted by a less simple-minded conversation model incorporating a formalization of speech acts (as, e.g.,in Moulin and Rousseau [22]). Moreover, the conversational manager maintains a simple form of turn-taking management, by assigning agents to take turns based on their personality traits. E.g., if James is an extrovert waiter, he would tend to start a conversation with a customer, which is formalized as

$$initiative(james,extrovert,take)$$

The environmental manager simulates the world that agents inhabit and updates its (shared) knowledge base with consequences of their actions. E.g., if the agent character Al got his beer in situation $s5$, this will be stored as

$$holds(al,has\_beer,s5)$$

The characteristics of the environment are encoded by a set of facts and rules. Situation calculus is used to describe and reason about change in the environment (Elkan [7]).

4. MENTAL MODELS

Each agent involved in the conversation is assumed to have its own mental model. A mental model may contain different kinds of entities, including world knowledge (beliefs, plans) and affective mental states (emotions, personality, moods, goals, attitudes). In this paper, we will concentrate on reasoning about affective states and social reasoning.

4.1 Reasoning about Emotion vs. Emotion Expression

It is widely accepted that animated agents expressing emotions are important to make the interaction with them more enjoyable and compelling for users (e.g., Lester *et al.* [17]). Emotional behavior can be conveyed through various channels, such as facial display, speech and body movement. The so-called ‘basic emotions’ approach (Ekman [6]) distills those emotions that have distinctive (facial) expressions associated with them: *fear*, *anger*, *sadness*, *happiness*, *disgust*, and *surprise*. Murray and Arnott [23] describe the vocal effects on

Emotion type ‘joy’: agent L is in a ‘joy’ state about state-of-affairs F with intensity δ in situation S **if**

> L wants F in S with desirability degree $\delta_{Des(F)}$
> **and** F holds in S **and** $\delta = \delta_{Des(F)}$.

Emotion type ‘angry-at’: agent $L1$ is angry at another agent $L2$ about action A with intensity δ in S **if**

> agent $L2$ performed action A prior to S
> **and** action A causes F to hold in S
> **and** agent $L1$ wants $\neg F$ with degree $\delta_{Des(\neg F)}$ in S
> **and** $L1$ considers A blameworthy with degree $\delta_{Acc(A)}$
> **and** $\delta = \log_2(2^{Des(\neg F)} + 2^{Acc(A)})$.

Figure 1: Specifications for *joy* and *angry-at.*

the basic emotions found in [6], e.g., if a speaker expresses the emotion ‘happiness’, his or her speech is typically faster, higher-pitched, and slightly louder.

Although a ‘basic emotions’ theory allows relating emotion to behavior (emotion expression), it cannot answer the question *why* an agent is in a certain emotional state. However, reasoning about emotions is considered equally important for presentation, pedagogical, and entertainment agents (e.g., André *et al.* [1], Johnson *et al.* [16], Rousseau and Hayes-Roth [28]). Many systems that reason about emotions, so-called *affective reasoners*, derive from the influential ‘cognitive appraisal for emotions’ model of Ortony, Clore, and Collins [25], also known as the OCC model (e.g., Elliott [8], O’Rorke and Ortony [24]). Here, emotions are seen as valenced reactions to events, agents’ actions, and objects, qualified by the agents’ goals (what the agent wants), standards (what the agent considers acceptable), and attitudes (what the agent considers appealing). The OCC model groups emotion types according to cognitive eliciting conditions. In total, twenty-two classes of eliciting conditions are identified and labeled by a word or phrase, such as ‘joy’, or ‘angry at’. We defined rules for a subset of the OCC emotion types: *joy*, *distress*, *hope*, *fear*, *happy for*, *sorry for*, *angry at*, *gloats*, and *resents* (see also O’Rorke and Ortony [24], Gratch [9]). In Fig. 1, the emotion types *joy* and *angry-at* are described.[1] The intensities of emotions are computed as follows (for all intensity degrees, $\delta \in \{1, 2, \ldots, 5\}$). In case of ‘joy’, we set $\delta = \delta_{Des(F)}$. For emotions, where intensities δ and γ have to be combined, such as in the specification of the ‘angry-at’ emotion, logarithmic combination is employed (Elliott [8]), $\delta = \log_2(2^\delta + 2^\gamma)$.

By example, let us explain the ‘angry at’ emotion type. Assume that a secretary is angry at her manager because she is refused to take a vacation. If the secretary has the goal to take vacation with desirability degree $\delta_{Des} = 3$, and considers the refusal as blameworthy with degree $\delta_{Acc} = 3$, then she will be angry at her manager with intensity degree $\delta = 4$. But how will she react to her manager? Presumably she will nod, showing that she understood the manager’s

[1] The specification of the ‘joy’ emotion is related to the specification of the ‘satisfaction’ emotion, whereby the latter one is prospect-based. An agent L is satisfied if a hoped-for state-of-affairs F holds, where L hopes for F if L wants F and anticipates F. See Gratch [9] for an in-depth treatment of computing intensities for prospect-based emotions.

answer, and try to convince the manager that she really needs some days off in a calm voice, with a rather neutral facial expression.

The secretary's behavior—*suppressing* the expression of her emotional state—can be explained in at least two ways. *First*, she might have personality traits that characterize her as very friendly. *Second*, and probably more important in this scenario, she might be aware of her social role as an employee which puts behavioral restrictions on her answer to the manager. Having said this, we should make explicit that we only consider deliberative forms of emotion (expression), as opposed to automatic 'hard-wired' processes of emotion expression (see Picard [26]).

Below, we provide a brief characterization of personality, and in the following we will try to explicate the impact of the social dimension on emotion expression in communication.

4.2 Personality

Personality traits are typically characterized by patterns of thought, attitude, and behavior that are permanent or at least change very slowly. Most importantly, believable agents should be *consistent* in their behavior (Rousseau and Hayes-Roth [28]). To keep things simple, we consider only two dimensions of personality, which seem crucial for social interaction. *Extroversion* refers to an agent's tendency to take action (e.g., being active, talkative). *Agreeableness* refers to an agent's disposition to be sympathetic (e.g., being friendly, good-natured). We assume numerical quantification of dimensions, with a value out of $\{-3, -2, -1, 1, 2, 3\}$. If an animated agent called 'James' is very outgoing and slightly unfriendly, it is formalized as

personality_type(james,extrovert,3,agreeable,-1)

As mentioned in Section 3, we consider the first dimension, extroversion, in the conversational manager: outgoing agents try to take turn in a conversation whenever possible, whereas introverted agents only respond when offered to take turn.

Moffat [20] points out the close relationship between personality and emotion, although they seem very different: emotions are short-lived and focused whereas personality is stable and global. He also considers *mood* which is rather short-lived (like emotion) and not focused (like personality). Later on, we will consider personality (or mood) to bias emotion expression given a certain emotional state. Consequently, an agent that is sorry for another agent and is friendly will express its emotion more intensely than an unfriendly agent.

5. SOCIAL FILTER PROGRAMS

Basically, a social filter program consists of a set of rules that encode qualifying conditions for emotion expression. The program acts as a 'filter' between the agent's affective state and its rendering in a social context, such as a conversation. Hence, we prefer to talk about social *filter* programs rather than control programs (Gratch [10]). We consider the agent's personality and the agent's social role as the most important emotion expression qualifying conditions.

5.1 Roles, Conventions, and Social Networks

A significant portion of human conversation takes place in a socio-organizational setting where participating agents have clearly defined *social roles*, such as sales person and customer, or teacher and student (Moulin [21]). Each role has associated behavioral constraints, i.e., responsibilities, rights, duties, prohibitions, and possibilities. Depending on its role, an agent has to obey communicative conventions (Lewis [18]). These conventions function as a regulatory for the agent's choice of verbal expressions in a given context. *Conventional practices* (i.e., behavioral constraints and communicative conventions) can be conceived as guidelines about socially appropriate behavior in a particular organizational setting. In this paper, we will focus on the choice of verbal and non-verbal behavior (emotion expression), depending on the agent's social role and personality.

Formally, in social or organizational groups roles are ordered according to a *power scale*, which defines the social power of an agent's role over other roles. For agents L_i and L_j, the power P of L_i over L_j is expressed as $P = p(L_i, L_j)$, where $P \in \{0, 1, 2, 3\}$. If $P = 0$, L_j considers itself as of the same rank as L_i. The *social network* is specified by the social roles and associated power relations. Walker *et al.* [29] also consider *social distance* between speaker and hearer to determine an appropriate linguistic style. Similarly, we use $D = d(L_i, L_j)$ to express the distance between two agents ($D \in \{0, 1, 2, 3\}$). Given values for power and distance, an agent L_i computes the (social) *threat* θ from agent L_j, by just adding the values, i.e.,

$$\theta = p(L_j, L_i) + d(L_i, L_j).$$

This is of course a very simple view of a social network but, as shown below, it already allows us to explain various phenomena in actual conversations. Observe that a zero value for threat can be interpreted in three ways: (i) there is no threat for an agent L, (ii) L chooses not to respect conventional practices, and (iii) L is not aware of any threat.

5.2 Social Filter Rules

In the following, we will give some examples of social filter rules. We assume that emotion expression (e.g., facial display or linguistic style) is determined by personal experience, background knowledge, and cultural norms (Walker *et al.* [29]), as well as the 'organizational culture' (Moulin [21]). Our rules are consistent with Brown and Levinson's theory of social interaction, as reported in [29].

If the conversational partner has more social power or distance is high (i.e., θ is high), the expression of 'negative' emotions is typically suppressed, resulting in 'neutralized' emotion expression (see Fig. 2). The first condition of the rule for emotion expression of 'anger' concerns the social context, the second condition the agent's personality (agreeableness), and the third accounts for the output of the affective reasoner, the emotional state. The intensity ϵ of emotion expression is computed as $\epsilon = \delta - (1 + \alpha + \theta)$. Consider the case of an agent that is very angry (i.e., $\delta = 5$), rather unfriendly (i.e., $\alpha = -2$), but considers the social threat as maximal (i.e., $\theta = 6$). Then $\epsilon = 0$, meaning that the angry emotion is completely suppressed. On the other hand, if $\theta = 0$, the agent's agreeableness dimension comes into force, resulting in $\epsilon = 6$ $(= 5 - (1 - 2))$. Since we assume five the maximal intensity level, greater values are cut off.

As shown in the second rule in Fig. 2, an agent might even express happiness about something which—*the agent believes*—distresses another agent. Observe that here, the agent has to reason about the emotions of another agent. Currently, we employ two mechanisms to model the ap-

Emotion expression 'anger': agent $L1$ displays expression 'anger' towards $L2$ with intensity ϵ **if**

the social threat for $L1$ from $L2$ is θ
and $L1$'s agreeableness has degree α
and $L1$ is angry at $L2$ with intensity δ
and $\epsilon = \delta - (1 + \alpha + \theta)$.

Emotion expression 'happiness': agent $L1$ displays expression 'happiness' towards $L2$ with intensity ϵ **if**

the social threat for $L1$ from $L2$ is θ
and $L1$'s agreeableness has degree α
and $L1$ is gloats at $L2$ with intensity δ
and $\epsilon = \delta - (1 + \alpha + \theta)$.

Emotion expression 'happiness': agent $L1$ displays expression 'happiness' towards $L2$ with intensity ϵ **if**

the social threat for $L1$ from $L2$ is θ
and $L1$'s agreeableness has degree α
and $L1$ is joyful with intensity δ
and $\epsilon = \delta - (\theta - \alpha)$.

Figure 2: Some examples of social filter rules.

praisal of another agent. If the observing agent has beliefs about the observed agent's mental states and their desirability, the agent infers the emotional state of the other agent by using its emotion rules. Else, the observing agent uses the other agent's perceived emotion, communicated via the *com_act/6* representation discussed in Section 3, to assess the other agent's emotion.

The third rule in Fig. 2 demonstrates the effect of personality and social context on 'positive' emotions. We compute the intensity of positive emotions as $\epsilon = \delta-(\theta-\alpha)$. As a consequence, the agent's unfriendliness or a high social threat will diminish the expression of positive emotions. E.g., if a very happy ($\delta = 5$) but rather unfriendly ($\alpha = -2$) agent communicates with a slightly distant agent (i.e., $\theta = 1$), the agent will express happiness with rather low intensity ($\epsilon = 2$).

Finally, notice an interesting consequence of our framework. Since we clearly distinguish between emotional state and expression of emotion, we may add another possibility of an agent's misinterpretation of other agents' behavior. *First*, an agent never has direct access to others mental states, it can only have (possibly false) beliefs about their mental (e.g., emotional) states. *Second*, our distinction allows that agents *cheat* in their behavior by expressing a misleading emotion. E.g., an agent may express a sad emotion, pretending to be in a distressed emotional state, although it is in a 'happy' state. This option is required for entertainment purposes, where 'levels of indirection' are beneficial.

5.3 Violations of Conventional Practices

Despite the fact that obedience to conventional practices is expected in real-world socio-organizational settings, violations of conventional practices occur, and in particular, they seem to be dramatically more interesting. What happens if a manager requests something from his or her secretary and the secretary refuses to follow the order? Consider the following conversation fragment:

Manager (to secretary): *I quickly need a copy of this.*
Secretary (to manager): *Sorry, I am busy right now.*

Here, the secretary violates conventional practices by ignoring the manager's indirectly formulated order. This situation typically triggers a negotiation process where the agent with higher social role makes his or her request more explicit, or the even directly refers to his or her role and associated rights, e.g., the power to request tasks from subordinates (Moulin [21]). From an emotion expression point of view, the manager agent will typically be in an 'angry at' (the secretary) emotional state and may express its anger emotion due to its higher rank on the power scale. The crucial belief in the manager agent's mental model is, e.g.,

blameworthy(manager,conventional_practice_violated,4)

i.e., managers (typically) consider it as blameworthy if their higher rank is ignored by employees. On the other hand, we might imagine a manager with friendly personality traits who is distressed about the secretary's behavior, and expresses the emotion 'sadness'. Another possibility is that the manager shows 'neutral' emotion expression, if manager and secretary are old friends and—according to the very organizational culture—typical conventional practices are not applicable.

6. ROLE-PLAYING ON THE WEB

Our interactive environment for English conversation training for Japanese speakers assumes that users (language students) would enjoy getting involved in a role-play with animated character agents, and thereby overcome their uneasiness to converse in a foreign language. Inspired by the examples of Rousseau and Hayes-Roth [28], we implemented an interactive theater (or drama) that offers the role of a customer in a virtual coffee shop. An interactive role-playing game employing animated agents, the *Wumpus Game*, is currently under development.

6.1 Implementation

The programmable interface of the Microsoft agent package [19] is used to run a virtual coffee shop session in a web browser (Internet Explorer 5). This choice put some serious restrictions from the outset: the characters available for this package have only a limited number of behaviors ('animations'), confining the realization of various emotions as well as some features of embodied conversational behavior (Cassell and Thórisson [4]). However, our goal is believablity on the level of adequate emotion expression rather than lifelikeness (in the sense of realistic behavior). The Microsoft Agent package provides controls to embed animated characters into a web page based JavaScript interface, and includes a voice recognizer and a text-to-speech engine. Prolog programs implement all reasoning related to conversation and environmental management and agents' mental models (affective and social reasoning). We use Jinni (Java Inference engine and Networked Interactor) to communicate between Prolog code and the Java objects that control the agents through JavaScript code (BinNet Corp. [3]).

In a role-playing session, the user can promote the development of the conversation by uttering one of a set of predefined sentences that are displayed on the screen. Unlike the setup of Hayes-Roth's 'Virtual Theater Project' [28], the

```
% emotion type 'angry at' in situation s1
holds(did(order_beer,customer),s1).
causes(order_beer,regulation_violated),s0).
blameworthy(james,order_beer,4).
wants(james,regulation_respected,3,s1).
% emotion expression 'anger' in situation s1
personality_type(james,extrovert,-2,agreeable,-3).
social_power(customer,james,0).
social_distance(james,customer,0).
% emotion type 'angry at' in situation s5
holds(did(refuse_vacation,manager),s5).
causes(refuse_vacation,no_vacation,s4).
blameworthy(james,refuse_vacation,3).
wants(james,get_vacation,5,s5).
% emotion expression 'neutral' in situation s5
social_power(manager,james,3).
social_distance(james,manager,2).
```

Figure 3: Some facts in the waiter agent's mental model for first example run.

user does not need an avatar in the play. Animated agents will respond by synthetic speech, facial display, and gestures. Verbal and non-verbal behavior is synthesized in the agent's mental model and interpreted in the browser. The parameters for speech output are set in accordance with the vocal effects associated with the basic emotions [6, 23]. Of course, the facial display of characters is limited to the predefined 'animations' from the Agent package (e.g., 'pleased', 'sad'). To some extent, we also implemented conversational behavior (Cassell and Thórisson [4]). E.g., the animations 'confused' (lifting shoulders) and 'don't-recognize' (put hand to ear) are used if the user's utterance is not recognized.

6.2 Example Runs

We will illustrate our system by showing some example runs. In the first example, the user takes the role of a (friendly) customer who interacts with an unfriendly waiter agent James, who himself interacts with a friendly manager agent as an employee. Fig. 3 describes some relevant facts stored in the waiter agent's brain. For the rule part, the reader is referred to Figures 1 and 2. The following is an annotated trace from our conversation system.

[s0] Customer: *I would like to drink a beer.* [User can also choose other beverages, and for each, he or she may select the linguistic style (polite, neutral, rude).]

[s1] James (to customer): *No way, this is a coffee shop.* [Considers it as blameworthy to be asked for alcohol and shows his anger. We assume that the waiter ignores the social threat from the customer.]

[s2] [The manager of the coffee shop appears.]

[s3] James (to manager): *Good afternoon, boss. May I take a day off tomorrow?* [Welcome gesture. Following conventional practices, the waiter is polite to his manager.]

[s4] Manager: *It will be a busy day. So I kindly ask you to come.* [Uses polite linguistic style in accordance with his personality traits.]

[s5] Waiter: *Ok, I will be there.* [Considers it as blameworthy to be denied a vacation and is angry. However, he is aware of the social threat and thus does not show his anger. Instead, he shows neutral emotion expression.]

The communicative act of the customer has the form

com_act(customer,james,order_beer,polite,neutral,s0)

Since the animated agents do not understand English, a library is used to associate the user's utterance with an 'effect', e.g., the regulations of the coffee shop are violated, and an evaluation of its linguistic style, such as polite, rude or neutral. Moreover, as an emotion (expression) recognition module is not part of our system, we set 'neutral' as the default value for user input (but see the work of Picard [26]). The waiter's answer is formalized as

com_act(james,customer,refuse_beer,rude,anger,s1)

Similarly, the library is employed to generate the syntactic form of the animated agent's response. As described in Section 3, the environmental manager simulates the environment. In this example, it includes the fact *act(manager, appears, s2)* which triggers the waiter's reaction in situation $s3$. In accordance with the contents of James' mental model and our rules for affective and social reasoning, the waiter agent expresses its anger towards the customer (user), but suppresses its anger towards its manager.

The second example run is a variation of the previous example where we assume an extrovert, friendly waiter who respects conventional practices towards customers but not towards his indifferent manager.

[s0] James: *Welcome to our coffee shop! May I take your order?* [Starts the conversation because of his extrovert personality.]

[s1] Customer: *Bring me beer, right away.* [User chooses rude linguistic style.]

[s2] James (to customer): *I am sorry but I am not allowed to serve alcoholic beverages here.* [Concludes that the customer is distressed and feels sorry for the customer.]

[s3] [The manager of the coffee shop appears.]

[s4] James (to manager): *Good to see you, boss. Tomorrow I want to take a day off.* [Performs welcome gesture.]

[s5] Manager: *Actually, I need you tomorrow. Thank you.* [Uses neutral linguistic style for his request.]

[s6] James: *Too bad for you. I will not be here.* [Waiter is angry as the manager refuses to give him a vacation. Since the waiter does not respect conventional practices, he expresses his anger and refuses to obey the manager's order.]

In situation $s3$, we assume James to believe that the customer wants a beer (urgently) and is distressed as a consequence of the waiter's refusal. James' agreeableness is responsible for feeling sorry about having to refuse the customer's order. However, the linguistic style of James' response is slightly lower in its politeness, since the customer approaches the waiter in a rude way (a rudimentary form of a *reciprocal feedback loop*).

There are limitless ways to vary social encounters between agents, even in the restricted context of a coffee shop environment. Consider, for instance, a situation where an unfriendly waiter shows rude behavior towards a customer, who turns out to be the waiter's new manager. A 'socially robust' waiter agent will show a form of 'behavior switching' (assuming that the agent respects conventional practices).

6.3 User Feedback

We conducted a rather preliminary experiment on the impact of agents featuring social role awareness. As in the example runs of the previous section, users would play the role of a customer in a coffee shop and interact with an animated agent portraying a waiter. The waiter agent interacts with a manager agent, a fellow waiter agent or a customer agent. Although our general goal is to employ the animated agents approach to language conversation training, the focus here is to show that (i) users can recognize whether the agents behave according to their social roles, and (ii) the animated agents' responses are believable. Five users were asked to rate the appropriateness of the agents' responses. Furthermore, we asked them whether they think the agents' reactions could occur in real-world situations.

As we expected, users could identify the social roles played by the animated agents, which are easily detected in the coffee shop environment. Users could also recognize when conventional practices have been violated. However, answers varied when asked for what they think went wrong in case of violation: including, that the agent is in a bad mood, or does not like the boss (or customer agent). When we run the experiment with agents that only reason about emotion and personality (i.e., without social role awareness), users would generally not consider those agents as 'unbelievable', but they expected to get hints regarding the motivation for the violation of conventional practices, and appropriate reactions from the other agents. Our (preliminary) findings reinforce the belief that an agent should show an overall consistency in its behavior in order to come across as believable and that social role awareness facilitates consistency.

7. DISCUSSION

Our work aims to account for an important feature of human-human communication, namely social role awareness, that seems to have strong influence on our ways of emotion expression and our behavior in general. Social role awareness is approached from the viewpoint of the believability of animated agents. It is shown that this feature of social interaction may explain phenomena such as suppressing (the expression) of emotions, as well as other forms of 'cheating' about an agent's emotion. As such, social role awareness can significantly contribute to the design of dramatically interesting characters (as in Hayes-Roth *et al.* [13] or Rousseau and Hayes-Roth [28]). More recently, sensitivity to socio-organizational contexts is also pointed out as a crucial issue in military training simulations (Gratch and Hill [11], Gratch [10]). Here, interesting conflicts can arise between an agent's goal (or self-interest) and role-specific duties imposed by orders from a commander agent.

Although we believe that social role awareness makes animated agents more 'socially robust', our approach suffers from several shortcomings. In the following paragraphs, we discuss work relevant to future refinements of our approach.

Social reasoning and planning. In the emotion model called *Émile*, Gratch [9] interleaves emotional reasoning with an explicit planning model. There are obvious advantages of considering an agent's plans, e.g., 'prospective' emotional states such as hope and fear assume reasoning about future situations and typically induce plan generation or the modification of current plans. Similarly, social reasoning would benefit from an explicit representation of plans. Often, an agent's choice of emotion expression depends on the state of its current plan: e.g., if an employee agent plans to get fired, its violation of conventional practices towards its manager should be seen in the light of this high-level goal, and not be considered as part of the agent's concept of its social role.

Social action. Besides appropriate emotion expression, other (possibly more important) behavioral constraints apply to socio-organizational settings. The role of an agent is associated with certain responsibilities, rights, duties, prohibitions, and decision power (Moulin [21]). A broader perspective of social agency will require explicit representations of those behavioral constraints, as well as formalisations of social concepts such as 'commitment'. In this respect we may heavily draw on well-established research work on multi-agent systems and distributed artificial intelligence (e.g., Jennings [15], Castelfranchi [5]).

Social communication. As mentioned throughout the paper, an obvious weakness of our approach is that we do not provide an explicit formalization of speech acts. Consequently, all of the dialogue contributions have to be carefully hand-crafted. In fact, we employed a simplified version of Moulin and Rousseau's conversation model [21, 22]. In [21], Moulin introduces a new notation for speech acts that is tailored to communication in socio-organizational settings, in particular *conversational schemas* that allow an agent to select speech acts in accordance with communicative conventions. In addition, we should consider the *linguistic style (LS) strategies* discussed by Walker *et al.* [29]. Those strategies determine semantic content, syntactic form and acoustical realization of a speech act, qualified by the social situation. Application of LS strategies supports social interactions that allow agents to maintain public face (i.e., autonomy and approval). If speaker and hearer have equal social rank, 'direct' strategies can be applied (e.g., "Bring me a beer!"). On the other extreme, if the rank distance is very large, 'off record' strategies are chosen (e.g., "Someone has not brought me a beer.").

8. CONCLUSION

In this paper, we propose to integrate social reasoning to mental models of animated agents, in addition to an affective reasoning component. The novel aspect of our work is that we explicate the social role of agents and associated constraints on emotion expression, which allows for enhanced believability of animated characters beyond reasoning about emotion and personality. We believe that considering the social dimension in animated agents approaches adds value for the following reasons:

- *Believability.* It may increase the illusion of life, which is often captured by emotion and personality only.
- *Social Communication.* By respecting an important feature of human conversation, it adds 'social robustness' to agent-human and inter-agent communication.
- *Explanatory power.* It explains the frequent mismatch between the output of emotional reasoning (the emotional state) and emotion expression.

We have described a web-based interactive drama scenario featuring animated agents as an entertaining testbed to experiment with new capabilities of agents. By considering the issues described in the discussion section, we hope to gain a better understanding of the social dimension in communication.

9. ACKNOWLEDGMENTS

We would like to thank the anonymous referees for their very helpful and detailed comments. The first author was supported by a grant from the Japan Society for the Promotion of Science (JSPS).

10. REFERENCES

[1] E. André, T. Rist, S. van Mulken, M. Klesen, and S. Baldes. The automated design of believable dialogue for animated presentation teams. In J. Cassell, S. Prevost, J. Sullivan, and E. Churchill, editors, *Embodied Conversational Agents*, pages 220–255. The MIT Press, 2000.

[2] J. Bates. The role of emotion in believable agents. *Communications of the ACM*, 37(7):122–125, 1994.

[3] BinNet Corp. *Jinni 2000: A high performance Java based Prolog for agent scripting, client-server and internet programming*, 2000.

[4] J. Cassell and K. R. Thórisson. The power of a nod and a glance: Envelope vs. emotional feedback in animated conversational agents. *Applied Artificial Intelligence*, 13(4–5):519–538, 1999.

[5] C. Castelfranchi. Modeling social action for AI agents. In *Proceedings 15th International Conference on Artificial Intelligence (IJCAI-97)*, pages 1567–1576, 1997. Invited Talk.

[6] P. Ekman. An argument for basic emotions. *Cognition and Emotion*, 6(3–4):169–200, 1992.

[7] C. Elkan. Incremental, approximate planning. In *Proceedings 8th National Conference on Artificial Intelligence (AAAI-90)*, pages 145–150, 1990.

[8] C. Elliott. *The Affective Reasoner. A process model of emotions in a multi-agent system.* PhD thesis, Northwestern University. Institute for the Learning Sciences, 1992.

[9] J. Gratch. Émile: Marshalling passions in training and education. In *Proceedings 4th International Conference on Autonomous Agents (Agents'2000)*, pages 325–332, 2000.

[10] J. Gratch. Socially situated planning. In K. Dautenhahn, editor, *Socially Intelligent Agents: The Human in the Loop (Papers from the 2000 AAAI Fall Symposium)*, pages 61–64, 2000.

[11] J. Gratch and R. W. Hill. Continuous planning and collaboration for command and control in joint synthetic battlespaces. In *Proceedings 8th Conference on Computer Generated Forces and Behavioral Representation*, 1999.

[12] A. Guye-Vuillème and D. Thalmann. Requirements for an architecture for believable social agents. In *Proceedings 4th International Conference on Autonomous Agents (Agents 2000)*, pages 48–49, 2000.

[13] B. Hayes-Roth, R. van Gent, and D. Huber. Acting in character. In R. Trappl and P. Petta, editors, *Creating Personalities for Synthetic Actors: Towards Autonomous Personality*. Springer Verlag, 1997.

[14] M. Ishizuka, T. Tsutsui, S. Saeyor, H. Dohi, Y. Zong, and H. Prendinger. MPML: A multimodal presentation markup language with character control functions. In *Proceedings Agents'2000 Workshop on Achieving Human-like Behavior in Interactive Animated Agents*, pages 50–54, 2000.

[15] N. R. Jennings. Commitments and conventions: The foundation of coordination in multi-agent systems. *The Knowledge Engineering Review*, 8(3):223–250, 1993.

[16] W. L. Johnson, J. W. Rickel, and J. C. Lester. Animated pedagogical agents: Face-to-face interaction in interactive learning environments. *International Journal of Artificial Intelligence in Education*, 2000.

[17] J. C. Lester, S. G. Towns, and S. G. FitzGerald. Achieving affective impact: Visual emotive communication in lifelike pedagogical agents. *International Journal of AI in Education*, 10(3–4):278–291, 1999.

[18] D. Lewis. *Convention. A Philosophical Study.* Harvard University Press, 1969.

[19] Microsoft. *Developing for Microsoft Agent.* Microsoft Press, 1998.

[20] D. Moffat. Personality parameters and programs. In R. Trappl and P. Petta, editors, *Creating Personalities for Synthetic Actors*, pages 120–165. Springer, 1997.

[21] B. Moulin. The social dimension of interactions in multiagent systems. In W. Wobcke, M. Pagnucco, and C. Zhang, editors, *Agents and Multi-Agent Systems, Formalisms, Methodologies and Applications*, pages 109–122. Springer Verlag, 1998.

[22] B. Moulin and D. Rousseau. An approach for modeling and simulating conversations. In D. Vanderveken and S. Kubo, editors, *Essays in Speech Act Theory.* John Benjamins, 2000.

[23] I. R. Murray and J. L. Arnott. Implementation and testing of a system for producing emotion-by-rule in synthetic speech. *Speech Communication*, 16:369–390, 1995.

[24] P. O'Rorke and A. Ortony. Explaining emotions. *Cognitive Science*, 18:283–323, 1994.

[25] A. Ortony, G. L. Clore, and A. Collins. *The Cognitive Structure of Emotions.* Cambridge University Press, 1988.

[26] R. W. Picard. *Affective Computing.* The MIT Press, 1997.

[27] H. Prendinger and M. Ishizuka. Carrying the role-playing metaphor to interactive learning environments. In *International Conference on Intelligent User Interfaces (IUI-2001)*, 2001.

[28] D. Rousseau and B. Hayes-Roth. A social-psychological model for synthetic actors. In *Proceedings 2nd International Conference on Autonomous Agents (Agents'98)*, pages 165–172, 1998.

[29] M. A. Walker, J. E. Cahn, and S. J. Whittacker. Improvising linguistic style: Social and affective bases for agent personality. In *Proceedings First International Conference on Autonomous Agents (Agents'97)*, 1997.