The performance of the dogs highlighted several factors.

- Locally initiated behavior must be appropriate and this is not always obvious until the system is understood. Only after the trials of the first three dogs had been completed was this understanding gained, making it possible to design behaviors which were truly appropriate.
- Appropriate behavior can drastically reduce operator interaction. The results obtained with follow1 and weave1 demonstrate marked improvements over their predecessors.
- Engineering effort expended on behaviors can easily have a negative impact and do not always yield a proportional return. The skirt behavior, which was added to align1 with the intention of improving its ability, increased the number of interactions required while requiring effort to develop.

Complete automation of a task is a time consuming and complex engineering exercise. To achieve an initial reduction in the level of operator interaction required is a relatively easy task. To reduce the level of interaction further requires the development of an appropriate behavioral repertoire which requires the development of a detailed understanding of the system. This not only includes the robot and the task, but the real environment in which the robot interacts. These results seem to imply a diminishing return in reduction of operator interaction for a given investment in engineering effort.

References

[1] M. S. Wilson and M. J. Neal, "Telerobotic sheepdogs: How useful is autonomous behavior?," in *Proc. 6th Int. Conf. Simul. Adaptive Behavior*. Cambridge, MA, 2000, pp. 125–134.
[2] P. Bhatia and M. Uchiyama, "A VR-human interface for assisting human input in path planning for telerobots," *Presence: Teleoperators Virtual Environ.*, vol. 8, no. 3, pp. 332–354, 1999.
[3] V. Lumelsky, "On human performance in telerobotics," *IEEE Trans. Syst., Man, Cybern.*, vol. 21, pp. 971–982, Sept./Oct. 1991.
[4] J. Matijevic, "The pathfinder mission to Mars—Autonomous navigation and the sojourner microrover," *Science*, vol. 280, no. 5362, pp. 454–455, 1998.
[5] K. Brady *et al.*, "A modular telerobotic architecture for waste retrieval and remediation," *Int. J. Robot. Res.*, vol. 17, no. 4, pp. 450–460, 1998.
[6] J. Trevelyan, "Simplifying robotics–A challenge for research," *Robot. Autonom. Syst.*, vol. 21, pp. 207–220, 1997.
[7] K. Taylor and J. P. Trevelyan, "A telerobot on the world-wide-web," in *Proc. Robots Austral. Ind. Conf.*, 1995, pp. 108–120.
[8] L. Conway, R. A. Volz, and M. W. Walker, "Teleautonomous systems: Projecting and coordinating intelligient action at a distance," *IEEE Trans. Robot. Automat.*, vol. 6, pp. 146–158, Apr. 1990.
[9] R. C. Arkin and T. Balch, *Artificial Intelligience and Mobile Robots: Case Studies of Successful Robot Systems*. Cambridge, MA: AAAI/MIT Press, 1998, pp. 277–296.
[10] R. Parasuraman, T. B. Sheridan, and C. D. Wickens, "A model for types and levels of human interaction with automation," *IEEE Trans. Syst., Man, Cybern. A*, vol. 30, pp. 286–297, May 2000.
[11] Parallax, "BASIC Stamp Manual, version 1.9," Parallax, Inc., Rocklin, CA, 1998.
[12] T. E. Colles, J. C. Hallam, and T. Smithers, "Control process language compiler user manual," Dept. Artif. Intell., Univ. Edinburgh, Edinburgh, U.K., DAI Tech. Paper TP17A, 1993.
[13] T. J. McCabe, "A complexity measure," *IEEE Trans. Softw. Eng.*, vol. SE-2, pp. 308–320, Apr. 1976.

Let's Talk! Socially Intelligent Agents for Language Conversation Training

Helmut Prendinger and Mitsuru Ishizuka

***Abstract*—This paper promotes socially intelligent animated agents for the pedagogical task of English conversation training for native speakers of Japanese. Since student–agent conversations are realized as role-playing interactions, strong requirements are imposed on the agents' affective and social abilities. As a novel feature, *social role awareness* is introduced to animated conversational agents, that are by now strong affective reasoners, but otherwise often lack the social competence observed in humans. In particular, humans may easily adjust their behavior depending on their respective role in a social setting, whereas their synthetic pendants tend to be driven mostly by emotions and personality. Our main contribution is the incorporation of a "social filter program" to mental models of animated agents. This program may qualify an agent's expression of its emotional state by the social context, thereby enhancing the agent's believability as a conversational partner. Our implemented system is web-based and demonstrates socially aware animated agents in a virtual coffee shop environment. An experiment with our conversation system shows that users consider socially aware agents as more natural than agents that violate conventional practices.**

***Index Terms*—Affective reasoning, animated agents, believability, emotion expression, social dimension in communication, social role awareness.**

I. Introduction and Motivation

We can recently notice a shift of interest from believable agents to a new generation of agents that are characterized as "socially intelligent" [1]. Starting with Bates' seminal work on *believable* agents in their "Oz project" [2], there have been continued efforts to give animated agents the illusion of life. By now, it is widely accepted that emotion expression and personality are key components of believable agents [3]. Moreover, Cassell *et al.* [4] provided convincing evidence of the importance of nonverbal "embodied" conversational behavior for believable agents. In addition to believability, *socially intelligent* agents are intended to perform more natural and robust in interactions that are explicitly embedded in a social context with associated social goals, such as establishing and maintaining social relationships. In this area, opinions regarding the core issues are more diverse, possibly because the field is highly interdisciplinary. As a guideline, we report on Dautenhahn's [5] four assumptions ("working hypotheses") underlying her work on social agents.

1) Intelligence is linked to a body.
2) The body is adapted to its embedding environment.
3) Intelligence can only be studied when considering the interaction between body and environment.
4) Intelligent agents are social entities.

She defines *social intelligence* generally as "…the individual's capability to develop and manage relationships between individualized, autobiographic agents which, by means of communication, build up shared social interaction structures which help to integrate and manage the individual's basic ("selfish") interests in relationship to the interests of the social system at the next higher level."

Manuscript received January 4, 2001; revised April 30, 2001. This work is supported by a JSPS Research Grant (1999–2003) for the Future Program.
The authors are with the Department of Information and Communication Engineering, School of Engineering, University of Tokyo, Tokyo, Japan (e-mail: helmut@miv.t.u-tokyo.ac.jp; ishizuka@miv.t.u-tokyo.ac.jp).
Publisher Item Identifier S 1083-4427(01)07732-3.

In this paper, we will take a particular stance on socially intelligent agents. While our primary interest is agents' believability as conversational partners, we will argue that purely emotion and personality-based agent behavior falls short of "social robustness" as it does not account for the social context in which a conversation is embedded, e.g., a sales agent might be convincing (believable) in the role of a salesperson, but not in the role of a colleague or employee. Humans, on the other hand, are always aware of the roles they play in a certain social setting and typically behave accordingly. This important feature of human–human communication will be called *social role awareness* and constitutes our contribution to the social intelligence of agents.

As an example of social role awareness, consider an unfriendly sales agent that is angry at the customer because of the customer's detailed questions about a product. The agent will presumably use polite language and hide its anger as it is aware of the role-specific behavioral restrictions applicable to a sales conversation. On the other hand, imagine the sales agent interacts with a colleague (agent). In this social setting, we would expect that the sales agent behaves according to its personality traits (in our case, unfriendly). A basic assumption of our work is that an agent's behavior cannot be generated (or understood) by considering personality, attitudes, and emotions alone but has to integrate social role awareness as an essential component of its design.

We recently started a project with the aim to employ animated agents for the pedagogical task of *language conversation training* [6]. Specifically, the animated agent approach is used to improve English conversation skills of native speakers of Japanese. Interactions between language students and agents are implemented as role-playing dramas and games. In order to come across as believable and interesting conversational partners, the agents have to show coherent emotional responses and social behavior. As opposed to genuinely educational software that features an animated tutor, we do not face the plausibility (or acceptability) problem identified by Lepper *et al.* [7]. In fact—adhering to the language training as role-playing metaphor—the very assumption (and hope) of our approach is that animated agents are perceived as game characters rather than language instructors.

The rest of this paper is organized as follows. The next section discusses mental states underlying an agent's social reasoning and models of emotion and emotion expression. In Section III, we introduce the social filter program that functions as a filter between an agent's affective state and the agent's emotion expression and a simple model of social feedback. In Section IV, we first explain our human–agent role-playing environment and the system architecture. Then, we describe a framework for modeling and simulating conversations. After that, we illustrate our system by example runs of an interaction session. Section V reports on the results of an empirical study that investigates users' response to socially (un)aware agents. Finally, Section V summarizes the paper.

II. Mental Models

Each agent involved in a conversation is assumed to have its own mental model. A mental model may contain different kinds of entities, including world knowledge (beliefs) and mental states such as emotions, personality traits, attitudes, and goals. In this section, we will focus on the core components underlying socially intelligent reasoning.

A. Emotional States

Many systems that reason about emotion, so-called *affective reasoners*, derive from the influential "cognitive appraisal for emotions" model of Ortony *et al.* [8], which is also known as the OCC model. Here, emotions are seen as valanced 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 OCT model groups emotion types according to cognitive eliciting conditions. In total, 22 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 OCT emotion types: *joy*, *distress*, *hope*, *fear*, *happy-for*, *sorry-for*, *angry-at*, *gloats-at*, and *resents*. In Fig. 1, the emotion types *joy* and *angry-at* are described. Emotional states have associated intensities $\delta \in \{1, 2, \ldots, 5\}$. When intensities of emotions have to be combined, logarithmic combination is employed [9].

> **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* 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 to degree $\delta_{Acc(A)}$
> **and** $\delta = \log_2(2^{Des(\neg F)} + 2^{Acc(A)})$.

Fig. 1. Specifications for *joy* and *angry-at*.

B. Personality

Personality traits are typically characterized by patterns of thought and behavior that are permanent or at least change very slowly [10]. Consequently, believable agents should be *consistent* in their behavior [3]. For simplicity, we consider only two dimensions of personality, which seem crucial for social interaction.

- *Extraversion* refers to an agent's tendency to take action: sociable, active, talkative, optimistic.
- *Agreeableness* refers to an agent's disposition to be sympathetic: good-natured, helpful, forgiving.

We assume numerical quantification of dimensions, with a value out of $\{-3, -2, -1, 1, 2, 3\}$. For instance, a value of -3 in the agreeableness dimension means that the agent is very unfriendly.

C. Roles, Conventional Practices, 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 [11]. Each role has associated conventional practices that function as a regulatory for the agent's choice of verbal and nonverbal expressions. Those practices can be conceived as guidelines about socially appropriate behavior in a particular organizational setting.

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.* [12] 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.,

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

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 $\epsilon = \delta - (\theta - \alpha)$.

Fig. 2. Some examples of social filter rules.

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.

1) L is aware that there is no threat.
2) L chooses not to respect conventional practices
3) L is not aware of any threat.

D. Emotion Expression

Emotional behavior can be conveyed through various channels, such as facial display (expression), speech, and body movement. The so-called "basic emotions" approach [13] distills those emotions that have distinctive (facial) expressions associated with them and seem to be universal: *fear, anger, sadness, happiness, disgust,* and *surprise*. In our work, it is of prime importance to clearly distinguish between emotional states and emotion expression. In order to avoid confusion and since there is only a limited number of comprehensive "emotion words," we use *slanted* when referring to basic emotions instead of *italics* for emotional states.

The "basic emotions" approach provides a useful list of emotions as an emotion family inventory for animated agents as it explicitly relates emotion to behavior. Besides facial expressions, Ekman also suggests other signals, such as speech and body movement to express emotions. Most importantly, Murray and Arnott [14] describe the vocal effects of Ekman's basic emotions. For example, if a speaker expresses *happiness*, then his or her speech is typically faster (or slower) and higher-pitched, whereas a speaker expressing *sadness* usually uses slow and lower pitched speech.

III. Social Filter Program

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. We consider the agent's personality and the agent's awareness of its social role as the most important emotion expression qualifying conditions.

By way of example, let us explain the *angry-at* emotion type. Assume that a student is angry at her instructor because she is treated in an unfair way, which she considers as blameworthy. How will she react to her instructor? Presumably she will nod, showing that she understood the instructor's point of view, and try to argue that the instructor's interpretation is wrong, in a calm voice with rather neutral facial expression. The student'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 a student which puts behavioral restrictions on her answer to the instructor.

A. 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 [12], as well as the "organizational culture" [11]. Consequently, it is human agents who determine the values of the social variables "social power" and "social distance." Our rules are consistent with Brown and Levinson's theory of social interaction, as reported in [12].

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 $\epsilon = \delta - (1 + \alpha + \theta)$. The equations we currently use for computing the intensity of emotion expression are not "objective" but seem to bear some plausibility. 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-at* 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 is 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. We employ two mechanisms to model the appraisal of another agent. If the observing agent has beliefs about the observed agent's mental states and their desirability (a rather strong assumption), the agent infers the emotional state of the other agent by using its emotion rules. Otherwise, the observing agent uses the other agent's perceived emotion, communicated via a simple interaction protocol, which will be discussed in Section IV.

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. For example, if a very *joyful* $(\delta = 5)$ 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. For example, an agent may express sadness, pretending to be in a *distress* emotional state, although it is in a *joy* state. This possibility is especially interesting for entertainment purposes, where "levels of indirection" are required.

B. Social Feedback

Social filter programs certainly support high social accuracy of agents' responses. However, we currently have no principled way to introduce *social feedback* mechanisms to our model, i.e., values for social power and distance are kept fixed during the interaction. However, when humans establish and maintain social relationships, e.g., the value of their social distance changes. This phenomenon is not reflected in our model.

Another interesting phenomenon of human–human communication are *reciprocal feedback loops* where, e.g., one agent's use of polite linguistic style results in another agent adapting its linguistic style and eventually adjusting its value of social distance. Our system supports a limited form of reciprocal feedback, whereby depending on the user's (or agent's) linguistic style, "intensity units" are added or subtracted to (from) the agreeableness degree. Hence, if the agent would give a cheerful answer with intensity degree $\epsilon = 3$, it might respond with degree 5 if asked politely and with degree 1 if asked in a rude way (given appropriate intensity values for the remaining mental concepts). A neutral question does not change the emotion expression intensity.

IV. Role-Playing Environment

Our interactive environment for English conversation training for Japanese speakers is based on the idea that users (language students) would enjoy getting involved in a role-play with animated, socially intelligent agents, and thereby overcome their uneasiness to converse in a foreign language. Although Japanese speakers have a good knowledge of English grammar and vocabulary, they mostly lack opportunities to actually listen to and speak English. Inspired by the Virtual Theater project [3], we implemented an interactive drama that offers the role of a customer in a virtual coffee shop. Our agents' task is to keep the user interested and motivated to interact with them repeatedly.

A. System Architecture

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. Animated agents will respond by synthetic speech, facial display, and gestures. We use the Microsoft Agent package [15] that provides controls to embed animated characters into a web page based JavaScript interface and includes a voice recognizer and a text-to-speech engine. The user's input is passed to a Java applet that communicates with three types of knowledge bases via the Jinni 2000 interface [16] (see Fig. 3).

- *Conversation and environment managers* maintain a model of the conversation and simulate the environment. They will be described in more detail below.
- *Affective and filter programs* are the (Prolog) reasoning engines for generating an agent's emotional state and determining the intensity of emotion expression.
- *Beliefs and behavior* store an agent's mental states (e.g., as in Fig. 4) and rules that encode agent behavior. Verbal and nonverbal behavior is synthesized from prescripted utterances and predefined animations.

The result of affective and social reasoning is passed back to the Java applet and interpreted in the browser via a JavaScript function. As shown in Fig. 3, the agent might react to the user by saying "Welcome to our Virtual Coffee Shop!" and thereby bow gracefully. Admittedly,

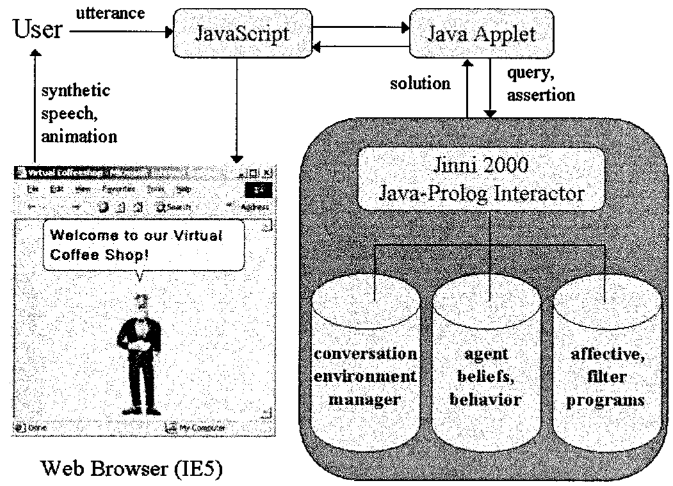

Fig. 3. Human–Agent interaction system architecture.

```
% 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).
```

Fig. 4. Some Prolog facts in the mental model of the waiter agent James. They are used in the first example run discussed in Section IV-C.

the choice of two–dimensional (2-D) cartoon-style agents as used in our system put some restrictions to the agents' reactions as the characters available for the Microsoft Agent package have only a limited number of predefined "animations" (about 50). However, our goal is believability on the level of socially adequate emotional response rather than lifelikeness (in the sense of realistic behavior). Also, McBreen and Jack [17] recently showed that 2-D agents are still acceptable to users, although three-dimensional (3-D) agents are clearly preferred.

B. Modeling and Simulating Conversations

In general, a conversation can be 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). Following Moulin and Rousseau [18], we 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.

TABLE I
CONVERSATION INVOLVING FRIENDLY CUSTOMER, UNFRIENDLY INTROVERT WAITER, AND HIS FRIENDLY MANAGER

Sit.	Speaker	Utterance	Annotation
s0	Customer	I would like a glass of beer.	User may select the linguistic style (polite, neutral, rude).
s1	Waiter	No way, this is a coffee shop.	The waiter agent considers it as blameworthy to be asked for alcohol and shows that he is *angry*. The agent ignores conventional practices, as the social distance between waiter and customer is high.
s2	Manager	Hello James!	The manager of the coffee shop appears.
s3	Waiter	Good afternoon. May I take a day off tomorrow?	Performs welcome gesture. Being aware of the social threat from his manager, the waiter uses polite linguistic style.
s4	Manager	It will be a busy day.	Manager implies that the waiter should not take a day off.
s5	Waiter	Ok, I will be here.	Considers it as blameworthy to be denied a vacation and is *angry*. However, the waiter is aware of the threat from his boss (agent) and thus suppresses his *angry* emotion.

TABLE II
CONVERSATION INVOLVING UNFRIENDLY CUSTOMER, FRIENDLY EXTROVERT WAITER, AND HIS FRIENDLY MANAGER

Sit.	Speaker	Utterance	Annotation
s0	Waiter	Welcome to our Coffee Shop!	Starts the conversation because of his extrovert personality.
s1	Customer	Bring me a beer, right away.	User chooses rude linguistic style.
s2	Waiter	I am sorry but it seems you are in the wrong place. We are not allowed to serve alcohol here.	Concludes that the customer is *distressed* and feels *sorry for* the customer. The intensity of the waiter's emotion expression is diminished by the fact that the customer's linguistic style is rude.
s3	Manager	Hello James!	The manager of the coffee shop appears.
s4	Waiter	Good to see you. Tomorrow I will take a day off.	Waves at manager in casual way.
s5	Manager	It will be a busy day.	
s6	Waiter	Too bad for you. I will not be here.	Waiter is angry as the manager refuses to allow a vacation. Since the waiter does not respect conventional practices towards the manager, the waiter expresses his *angry* emotion and refuses to obey the manager's order.

- 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 [11].

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?" According to our simple interaction protocol, the corresponding *communicative act* has the form

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 [18], we assume that 1) a conversation is governed by a *conversational manager* that maintains a model of the conversation and 2) 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 [18]. Moreover, the conversational manager maintains a simple form of turn-taking management by assigning agents to take turns based on their personality traits. For example, 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. For example, if the agent character Al got his beer in situation *5*, 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.

C. Example Runs

We will illustrate our system by showing two example runs. In the first example run, the user takes the role of a (friendly) customer who interacts with an unfriendly, introvert waiter agent (James) that interacts with a friendly manager agent as an employee. Table I shows the annotated trace from the interaction. Fig. 4 displays some of James' beliefs that have actually been used in the first example run.

The second example run is a variation of the previous example where we assume a friendly, extrovert waiter agent that is aware of conventional practices toward customers but not toward his friendly manager (see Table II).

V. EVALUATION

We conducted a small experiment on the impact of animated agents featuring social role awareness. As in the example runs above, participants would play the role of a customer in a virtual coffee shop and

TABLE III
MEAN SCORES FOR QUESTIONS ABOUT INTERACTION EXPERIENCE: RATINGS RANGE FROM 1 (DISAGREEMENT) TO 7 (AGREEMENT)

Question	Unfriendly Waiter (C1)	Friendly Waiter (C2)
James natural to user	3.00	6.00
James natural to others	4.88	5.50
James in real life, movie	5.00	4.63
James has good mood	2.25	2.25
James is agreeable	2.38	4.75
James likes his job	1.63	2.63

interact with an animated agent portraying a waiter. The waiter agent interacts with a manager agent and another customer agent that turns out to be an old acquaintance of the waiter. In the experiment, participants promoted the conversation by simply clicking a radio button next to the conversational contribution (that appeared in a separated window) instead of using the speech recognizer, so that they would not be distracted from the agents' reactions.

Sixteen participants, all students from the University of Tokyo, were randomly assigned to interact with one of two different versions of the system (eight subjects each). The two versions were identical except for the following features.

- In the **unfriendly waiter** version (C1), the waiter agent (James) responded to the user in a rude way, but changed to friendly behavior when interacting with his manager and the other customer (an old friend).
- In the **friendly waiter** version (C2), James displayed polite behavior to the user but disobeyed the manager's order and turned down his old friend.

After a 3-min interaction session, subjects were asked to fill out a questionnaire evaluating the naturalness (appropriateness) of James' behavior. We hypothesized the following outcome of the experiment.

- In the **unfriendly waiter** version (C1), subjects would rate James' behavior as unnatural toward themselves (as customers) but natural toward the other agents (manager, friend). Moreover, they would think that, in general, James has an unfriendly (disagreeable) personality.
- In the **friendly waiter** version (C2), on the other hand, subjects would consider James' behavior natural toward themselves but inappropriate toward the other agents and would find James' personality friendly.

T-tests (assuming unequal variances) on the data in Table III showed that subjects considered James' behavior significantly more natural (appropriate) in the C2 version than in the C1 version ($t = -4.4$; $p = .0011$). Concerning James' behavior toward the other agents, however, the experiment revealed the opposite of what we expected. Subjects considered James' behavior less natural in the C1 version (mean $= 4.88$) than in the C2 version (mean $= 5.5$). A possible reason is that although James ignored conventional practices toward the manager and the old friend in the C2 version, its behavior could still be considered as kidding. Another reason might be that due to the short interaction time, subjects could not figure out the personality of, e.g., James' manager. Consequently, if they assumed that James' manager is a very relaxed person, James' behavior could still be seen as appropriate.

We also asked subjects whether they could imagine meeting a waiter like James in a real coffee shop or as an actor in a movie. For both versions, subjects tended to agree, although less strongly than we expected (C1: mean $= 5$; variance $= 2.57$, C2: mean $= 4.63$; variance $= 3.41$). However, since James reacted consistently friendly/unfriendly toward each other agent, his behavior was still considered as believable.

Regarding James' personality, subjects found him significantly more agreeable in the C2 version than in the C1 version ($t = -3.5$; $p = .0019$). This result is interesting since in both versions, James shows (un)friendly behavior about half of the total interaction time. It supports our claim that behavior motivated by a social role, such as James' friendly behavior toward the manager in the C1 version, is conceived as part of the agent's social role and not his personality. Moreover, subjects considered the waiter's appreciation for his job significantly higher when James was friendly to the user than when he was friendly to his manager or friend ($t = -2.18$; $p = .0269$).

As to James' mood, we did not find any difference between the two versions (C1: mean $= 2.25$; variance $= 0.78$, C2: mean $= 2.25$; variance $= 0.22$). For the C2 version, this result shows that subjects clearly differentiate between personality and mood. On the other hand, Moffat's [10] work seems to imply that for sufficiently short time periods, it is hard to distinguish whether an agent's behavior is motivated by its mood or its personality.

VI. CONCLUSION

This paper 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 characters as socially intelligent 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 indirection in agents' affective behavior. As such, social role awareness can significantly contribute to the design of socially robust and dramatically interesting characters [3]. Our empirical study supports the hypothesis that users recognize and attribute socially aware behavior.

So far, we focused on the rather static aspects of socially intelligent behavior. This is clearly insufficient if we consider users that interact with animated agents repeatedly and thereby build up and maintain social relationships. We plan to extend our research by integrating a more elaborate mechanism of social dynamics to the agent's mental model. In particular, we are interested in changes (turns) of social parameters as a consequence of user–agent and agent–agent interaction, which is indispensable for dramatic action.

ACKNOWLEDGMENT

The authors would like to thank the reviewers for their sincere and valuable comments and suggestions.

REFERENCES

[1] K. Dautenhahn, Ed., "Socially intelligent agents: The human in the loop (Papers from the 2000 AAAI fall symposium)," AAAI Press, Tech. Rep., 2000.
[2] J. Bates, "The role of emotion in believable agents," *Commun. ACM*, vol. 37, no. 7, pp. 122–125, 1994.
[3] D. Rousseau and B. Hayes-Roth, "A social–psychological model for synthetic actors," in *Proc. 2nd Int. Conf. Autonom. Agents*, 1998, pp. 165–172.
[4] J. Cassell and K. R. Thórisson, "The power of a nod and a glance: envelope vs. emotional feedback in animated conversational agents," *Appl. Artif. Intell.*, vol. 13, no. 4–5, pp. 519–538, 1999.
[5] K. Dautenhahn, "Embodiment and interaction in socially intelligent life-like agents," in *Computation for Metaphors, Analogy and Agent*, C. L. Nehaniv, Ed., 1999, vol. LNAI 1562, pp. 102–142.
[6] H. Prendinger and M. Ishizuka, "Social role awareness in animated agents," in *Proc. 5th Int. Conf. Autonom. Agents*, 2001.

[7] M. R. Lepper *et al.*, “Motivational techniques of expert human tutors: lessons for the design of computer-based tutors,” in *Computers as Cognitive Tools*, S. P. Lajoie and S. J. Derry, Eds. New York: Lawrence Erlbaum, 1993, pp. 75–105.

[8] A. Ortony, G. L. Clore, and A. Collins, *The Cognitive Structure of Emotions*. Cambridge, U.K.: Cambridge Univ. Press, 1988.

[9] C. Elliott, “The affective reasoner. A process model of emotions in a multi-agent system,” Ph.D. dissertation, Northwestern Univ., Inst. Learning Sci., Evanston, IL, 1992.

[10] D. Moffat, “Personality parameters and programs,” in *Creating Personalities for Synthetic Actors*, R. Trappl and P. Petta, Eds., 1997, pp. 120–165.

[11] B. Moulin, “The social dimension of interactions in multiagent systems,” in *Agents and Multi-Agent Systems, Formalisms, Methodologies and Applications*, W. Wobcke, M. Pagnucco, and C. Zhang, Eds. New York: Springer-Verlag, 1998, pp. 109–122.

[12] M. A. Walker *et al.*, “Improvising linguistic style: social and affective bases for agent personality,” in *Proc. First Int. Conf. Autonom. Agents*, 1997.

[13] P. Ekman, “An argument for basic emotions,” *Cognit. Emotion*, vol. 6, no. 3–4, pp. 169–200, 1992.

[14] I. R. Murray and J. L. Arnott, “Implementation and testing of a system for producing emotion-by-rule in synthetic speech,” *Speech Commun.*, vol. 16, pp. 369–390, 1995.

[15] Microsoft, Inc., *Developing For Microsoft Agent*. Redmond, WA: Microsoft, 1998.

[16] *Jinni 2000: A High Performance Java Based Prolog for Agent Scripting, Client-Server and Internet Programming*, Bin Net Corp.,, 2000.

[17] H. McBreen and M. Jack, “Evaluating humanoid synthetic agents in E-retail applications,” *IEEE Trans. Syst., Man, Cybern A. Special Issue Socially Intelligent Agents*, vol. 31, pp. 394–405, Sept. 2001.

[18] B. Moulin and D. Rousseau, “An approach for modeling and simulating conversations,” in *Essays in Speech Act Theory*, D. Verveken and S. Kubo, Eds. Amsterdam, The Netherlands: John Benjamins, 2000.