Evolving Social Relationships with Animate Characters

Helmut Prendinger and Mitsuru Ishizuka
Department of Information and Communication Engineering
School of Engineering, University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
{helmut,ishizuka}@miv.t.u-tokyo.ac.jp

Abstract

In order to come across as life-like or animate, characters must change their attitudes ('evolve') as a consequence of their affective interaction histories with other agents. This paper describes a model of attitude and familiarity change that allows to model simple forms of evolving social relationships. We illustrate our model by means of two web-based interaction scenarios that feature animated characters in the role of comics-style actors.

1 Introduction and Motivation

In their work on 'animate characters', Hayes-Roth and Doyle (1998) describe how to endow agents with life and personality, so that users will enjoy interacting with them. We follow their naming convention, and use *animate* rather than *animated* in order to emphasize the importance of an agent's 'inner life' (or simply, character) rather than impressive graphics for its believability and life-likeness. But even if it is true that character comes first, there is vast evidence that an agent's embodiment may significantly contribute to the likeability of agents to users (McBreen et al., 2000). However, in applications of web-based characters or even characters for mobile devices restrictions on memory and processing power may prevent the use of highly expressive animations (Wahlster et al., 2001). In those settings, the agent's believability as an *animate* character will be a key concern.

Currently, the focus of most work in the field of embodied characters is on emotion modeling and expression (André et al., 1999; Gratch, 2000; Pelachaud and Poggi, 2001; Canamero and Fredslund, 2001). Emotion modeling basically describes the appraisal process, i.e., how events are evaluated as to their emotional significance for the agent. Often, characters are also endowed with a model of personality that gives their (re)actions consistency across different situations. Emotion expression, on the other hand, relates to various communicative modalities in which emotions are expressed, such as facial display, speech, and body posture. Characters appear animate because their embodied reactions to other agents or the user are life-like (believable). However, it seems to us that there is more to being animate than visually expressing an emotion. In particular, an agent may alter its emotional reactions over time, depending on the 'affective interaction history' with another agent. If some interlocutor triggers mostly positive (negative) emotions in the agent, it might change its attitude toward the interlocutor and be biased to appraise the interlocutor's future actions in a more positive (negative) way (Ortony, 1991).

In this paper, we focus on aspects of characters' internal lives that seem relevant for developing social relationships. More specifically, we try to model the change of attitudes and familiarity assessment that underly the appraisal process of attitude-based emotions such as *attraction* and *aversion*. In short, not only do attitudes contribute to the elicitation of emotions—but induced emotions may also change a character's affective state, in particular, its attitudes and familiarity toward another agent. This is important for the design of characters that are intended to be perceived as life-like or animate over extended periods of time, such as pet-type characters.

The next section briefly reviews related work. Section 3 first sketches an influential appraisal theory, and then develops a simple model of attitude and familiarity change. In Section 4, web-based interaction scenarios featuring life-like characters are introduced that implement our approach. Section 5 concludes the paper.

2 Related Work

The problem of modeling evolving social relationships has been addressed by various researchers from different disciplines. The biology-inspired area of 'Socially Intelligent Agents' aims to describe the mechanisms underlying the process of establishing and maintaining social relationships between human agents and artificial agents (Dautenhahn, 1998). Cañamero (2001) offers a good overview of issues relating to emotional agents in social interactions. Interestingly, she hypothesizes that agents will be more acceptable to humans if they "[...] reflect the evolution of emotional interactions over time—i.e. their 'history'." (p. 23). Breazeal and Velásquez (1998) model

a situation where Kismet, a robot infant, is engaged in social learning guided by a human caretaker. Cassell and Bickmore (2001) investigate the role of social language in social relationships. Specifically, they describe trust-promoting strategies between agents and humans, e.g., by the use of small talk.

3 Evolving Social Relationships

> "[...] I shall try to make a case for the claim that in addition to values being an important source of emotions, emotions are an important source of values and, more specifically, they can be the source of value in schemas." (Ortony, 1991, p. 341)

We will take Ortony's statement as a starting point for our model of evolving relationships and focus on one interpretation of values—as positive or negative attitudes toward an agent or object (liking and disliking). In the following, we will first briefly report on an interesting subset of the emotion model of Ortony *et al.* (1988), the so-called OCC model, and then describe how social relationships might change based on emotions elicited during social interaction between agents.

3.1 Modeling Emotions

According to the OCC model, emotions—or more precisely, emotion types—are simply classes of emotion eliciting conditions such as the agent's beliefs, goals, standards, and attitudes. An interesting set of emotions is crucially dependent on the agent's attitudes. The agent might be *happy for* its interlocutor if the agent likes the interlocutor and experiences *joy* over a state of affairs that it presumes to be desirable for the interlocutor. Otherwise, if the agent dislikes the interlocutor, it might *resent* its interlocutor for the same reason. A similar symmetry can be found with the *sorry for* and *gloat* emotion types. Let us now introduce some terminology by giving the specification of two other attitude-based emotion types, *attraction* and *aversion* (Ortony et al., 1988).

- A (locutor-)agent $L1$ is *attracted to* agent $L2$ with intensity δ *if* (i) $L1$ likes $L2$ with 'appealingness' degree δ_{App} and (ii) $L1$ is familiar to $L2$ with degree δ_F.
- An agent $L1$ has *aversion against* agent $L2$ with intensity δ *if* (i) $L1$ dislikes $L2$ with 'non-appealingness' degree δ_{NApp} and (ii) $L1$ is familiar to $L2$ with degree δ_F.

For simplicity, all intensity values range over $\{0, \ldots, 5\}$, with zero and five the lower and upper bounds, respectively, and are combined by means of logarithmic combination, e.g. $\delta = \log_2\left(2^{\delta_{App}} + 2^{\delta_F}\right)$ (see Prendinger and Ishizuka (2001) for more explanation).

Our interest here is the 'opposite direction' where emotions influence our attitudes toward an interlocutor, e.g., when we (suddenly) dislike the interlocutor because we got angry with her or him, or we like an interlocutor since she or he elicited mainly positive emotions in us Ortony (1991). We will also discuss an agent's change of familiarity (social distance) due to emotions elicited in the agent Pautler and Quilici (1998).

3.2 A Model of Attitude Change

Ortony (1991) suggests the notion of *(signed) summary record* to capture our attitude toward or dispositional (dis)-liking of another person. This record stores the sign of emotions (i.e., positive or negative) that were induced in the agent by an interlocutor together with emotions' associated intensities. For instance, if the interlocutor elicits *distress* with intensity 2, *anger* with intensity 1, and *joy* with intensity 5, the summary record of the agent will contain two values, a negatively signed value of 1 (3 divided by the number of situations), and a positively signed one of 1.67. In order to compute the current intensity of an agent's (dis)liking, we simply compare the (scaled) sum of intensities of elicited positive and negative emotions (δ^σ, $\sigma \in \{+, -\}$), starting in situation s_0, the situation when the interaction starts. We will only consider the 'winning' emotional state δ_w, i.e., the most dominant emotion, although in general, multiple emotions may be elicited in each situation. If no emotion of one sign is elicited in a situation, it is set to zero.

$$\delta^\sigma(s_n) = \frac{\sum_{i=0}^{n} \delta_w^\sigma(s_i)}{n+1}$$

Positive values for the difference $\delta^+ - \delta^-$ indicate an agent's liking of an interlocutor and negative ones indicate disliking in a certain situation.[1] For simplicity, we assume perfect memory of elicited emotions, i.e., the intensity of past (winning) emotions does not decay. If the interlocutor's recent behavior is mostly 'consistent' with the agent's past experience (i.e., both have same sign), it is reasonable to update the overall intensity of the agent's attitude according to the equation above.

Ortony (1991) also considers the more interesting case where an interlocutor the agent likes as a consequence of consistent reinforcement (suddenly) induces a high-intensity emotion of the opposite sign, e.g., by making the agent very angry. He suggests three types of reactions: (i) the agent is uncertain how to construct the summary record value; (ii) the agent updates the summary value by giving a greater weight to the inconsistent information; (iii) the agent ignores the inconsistent information in the construction of a summary value. Since there is plenty of evidence that recency of the interlocutor's inconsistent

[1] Situations are actually parameterized by the agent and interlocutor, referring to time points when emotions are elicited in an agent L as a result of communicating with interlocutor I.

behavior plays a significant role in determining an attitude (Anderson, 1965), we will focus on the second type of reaction. Although the notion of 'recency' could be generalized to m latest elicited emotions, we simply refer to the very latest emotion. Here, the update rule reads as follows (parameters are omitted).

$$\delta(s_n) = \delta^{\sigma}(s_{n-1}) \times \omega_h \mp \delta_w^{\overline{\sigma}}(s_n) \times \omega_r$$

The weights ω_h and ω_r denote the weights we apply to historical and recent information, respectively. ω_h and ω_r take values from the interval $[0, 1]$ and $\omega_h + \omega_r = 1$. A greater weight of recent information is reflected by using a greater value for ω_r. By way of example, let us assume that the agent likes its interlocutor with degree 3 and then gets angry at the interlocutor with intensity 5. The new value might be computed as $3 \times 0.25 - 5 \times 0.75$, resulting in a disliking value of 3.

The crucial question is how the obtained (dis)liking value affects future interactions with the interlocutor. We consider two interpretations:

- *Momentary (dis)liking.* The new value is active for the current situation and then enters the summary record.
- *Essential (dis)liking.* The new value replaces the summary record.

An instance of momentary liking are reciprocal feedback loops where a disagreeable interlocutor's (temporary) friendliness lets the agent adopt friendly behavior due to the elicitation of a positive emotion. Essential (dis)liking typically happens when the agent finds out something very positive (negative) about the interlocutor that is crucial for its model of the interlocutor.

It is interesting to observe that the way an agent deals with inconsistent information allows to make assumptions about its personality traits along the disagreeable–agreeable dimension. For instance, if the agent's attitude changes to essential disliking if made very angry once, it might be called unforgiving. Furthermore, a subtle interaction might exist between an agent's option for momentary or essential (dis)liking and the familiarity with the interlocutor. It can be argued that the most dramatic changes happen in recent evolving relationships, whereas agents familiar with each other rather experience momentary attitude changes.

3.3 A Simple Model of Familiarity Change

We take the notion of 'familiarity' to mean the social distance an agent has toward an interlocutor (Brown and Levinson, 1987). Per definition, we are close to family and friends, and distant to strangers. The problem we want to address in the following is how change in social distance might be captured computationally.

Pautler and Quilici (1998) investigate a special form of speech acts, called 'social perlocutions', that may change the interlocutor's relationship with the agent. They argue that positive emotions elicited in the interlocutor contribute to improving the interlocutor's social relationship with the agent. Although they do not explicitly discuss relationships in terms of familiarity, we belief that this interpretation is justifiable.

Our concept of familiarity degree δ_F considers the number and intensity of induced positive emotions δ^+.

$$\begin{aligned} \delta_F(s_0) &= 0 \text{ or pre-set to some value} \\ \delta_F(s_n) &= \delta_F(s_{n-1}) + \frac{\delta_w^+(s_n)}{\pi} \end{aligned}$$

If a negative emotion is elicited, $\delta_F(s_n) = \delta_F(s_{n-1})$. π is a factor that determines how rapid a character gets familiar with another agent. Unlike a character's (dis)liking, familiarity increases monotonically, i.e., once characters are socially close, they cannot subsequently get unfamiliar. Currently, our notion of familiarity is based on the (severe) simplifying assumption that emotions are taken as the only familiarity changing factor. Cassell and Bickmore (2001), on the other hand, consider the variety and depth of topics covered by conversing agents.

4 Web-based Interaction Scenarios

We will illustrate our model by means of two web-based scenarios that feature animated characters. In the first interaction scenario the user can play the "Black Jack" game in a virtual casino, and is guided by the animated advisor "Genie". The user may either follow or not follow Genie's advice, and independently of that, win or loose a game. Our goal is to show how the user's behavior as well as the (affective) interaction history determine Genie's reaction to the user's won or lost game. Genie appears animate since he may not express *joy* over a won game if the user repeatedly refuses to follow his advice.

The second interaction scenario involves a user controlled avatar who communicates with an angel-style character, by trying to figure out her wishes. Based on a very sparse emotion model, the aim of this scenario is to demonstrate how increased (decreased) liking and familiarity may 'switch' the angel character's affective reactions to *attraction* (*aversion*) toward the user's avatar. The character may appear animate as 'secondary' emotions such as attraction only occur as a result of a longer-term interactions.

In both scenarios, the Microsoft Agent package (Microsoft, 1998) is used to embed characters into a web page based JavaScript interface. An XML-based language called MPML (Multimodal Presentation Markup Language) is employed to specify sequential and parallel behavior of multiple characters (Descamps et al., 2001). For autonomous character control, MPML provides an interface to SCREAM, a tool that allows to script a character's mental state (including goals and personality traits) and

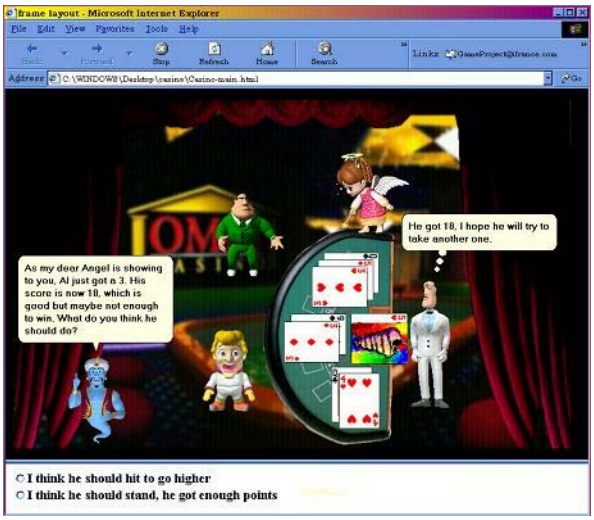

Figure 1: Casino Scenario.

mental processes such as emotion management and emotion regulation (Prendinger and Ishizuka, 2002). All characters are connected to a TTS (Text-to-Speech) engine and may express emotions by triggering pre-defined 2D animation sequences.

4.1 Playing Black Jack in a Virtual Casino

We will now watch the user playing five games of Black Jack and thereby demonstrate how Genie's mental make-up as well as the (affective) interaction history determine his behavior. Fig. 1 depicts the situation where the advisor Genie practices Black Jack with the user by commenting the game of character "Al" (Genie is the character at the bottom-left of the Internet Explorer window, and Al is the male character to the right of the dealer).

The following character profile is used (details about used intensities are omitted).

- *Genie's personality*. Agreeable and extrovert. His friendliness implies that negative emotions decay quickly whereas positive emotions remain active for a longer time.
- *Genie's goals*. He wants that the user wins with low intensity, and that the user follows his advice with high intensity.
- *Genie's social relations*. He is socially close to the user and initially slightly likes the user.

For expository reasons, we let the user *never* follow Genie's advice.

First game (user looses). Genie's winning emotional state is *distress* with high intensity, because the user did not follow his advice. However, he displays *distress* with low intensity as his agreeable personality effects a decrease in the intensity of negative emotion expression.

Second game (user looses). Genie is *sorry for* the user with high intensity, since positive ('sorry for' the user's lost game) emotions sum up (and decay slowly), which leads to an increase in Genie's liking of the user. His personality traits let him express the emotion with even higher intensity.

Third game (user looses). Genie *gloats* over the user's lost game, because at that point, the negative emotions dominate the positive ones as a consequence of the user's repeated refusal to follow Genie's advice. Hence Genie's attitude changes to slightly disliking the user which lets him experience *joy* over the user's *distress* (*gloat* with high intensity). Again, Genie's friendly personality decreases the intensity of the expressed emotion.

Fourth game (user wins). Genie's emotional state is *bad mood* with high intensity, slightly more than his *happy for* emotion (as the user wins the game this time). Here an overall, unspecific affective state (mood) is expressed with low intensity, rather than a specific emotion.

Fifth game (user wins). Genie's dominant emotion is *resent* with high intensity, because he slightly dislikes the user and consequently is feels *distress* that the user won although she or he ignored his advice. Genie expresses his emotion with reduced intensity.

An exhaustive exploration of all possible interaction patterns in the described game scenario reveals that Genie's reactions are conform at the beginning games and show more variety in the subsequent games. This can be explained by the development of Genie's attitude toward the user, depending on whether the user follows or refuses to follow Genie's advice. In effect, Genie's attitude decides, e.g., whether he is *sorry for* or *resents* the user's lost game. However, in accordance with Genie's agreeableness, his emotional reactions are mostly positive.

4.2 Interacting with Little Akko

Borrowing the idea from Fujio Akatsuka's manga series (Japanese comics) "Akko-chan's Got a Secret!", a character called 'Little Akko' (Akko-chan) plays the heroine of stories for kids. Little Akko has the power to be transformed into any person upon telling her wish to a magic mirror. By this magic, she has the power to solve many problems and even make other people happy. Figures 2 and 3 show her transformed into Little Chika, a girl whom her brother Kankichi likes. Social relationships in this comics book typically evolve in a quick and direct way and hence the stories lend themselves to easy testing of our model. We started to experiment with attitude and familiarity change based on a small set of emotion types: *joy*, *distress*, *attraction*, and *aversion*.

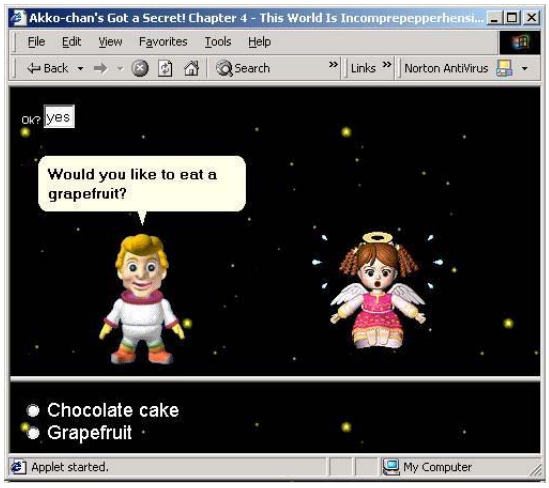

Figure 2: Angel is distressed toward Space-boy about getting a grapefruit.

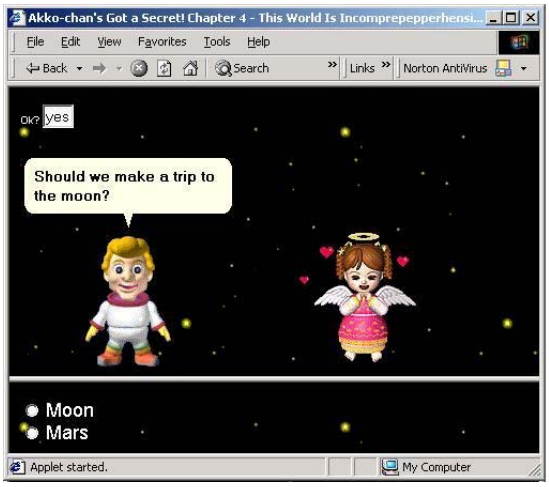

Figure 3: Angel is attracted to Space-boy after being offered a trip to the moon.

Currently, the interaction setting is fairly simple. The user can communicate with the "Angel" character (Little Akko transformed to Little Chika) by controlling an avatar, the "Space-boy" character in the role of Kankichi. By offering Little Chika items she likes, the user may increase her positive attitude and familiarity, otherwise her liking level for Kankichi goes down. Consider the following conversation.

- User may select "Strawberry milk" or "Lemon tea"
  Space-boy: Would you like to drink strawberry milk?
  Angel: Great! I like this drink.
- User may select "Chocolate cake" or "Grapefruit"
  Space-boy: Would you like to eat a grapefruit?
  Angel: No! I do not like that.
- User may select "Calculate" or "Hide and Seek"
  Space-boy: Do you want to play the Calculate Game?
  Angel: I really like that game!
- User may select "Rice" or "Noodles"
  Space-boy: Would you like to eat some rice?
  Angel: Yes! That is what I like!
- User may select "Moon" or "Mars"
  Space-boy: Should we make a trip to the moon?
  Angel: I enjoy being with you!

When Angel gets strawberry milk, she expresses *joy* as one of her goals is satisfied. After being offered a grapefruit, she shows her *distress* since she does not want this kind of dessert (see Fig. 2). However, in the conversation above, the user happens to repeatedly select items the Angel likes, which has two kinds of effects. Both the Angel's liking value toward the Space-boy and the familiarity level increase, and hence add to the intensity of the Angel's *attraction* toward the Space-boy (familiarity was incremented by 0.2 per elicited positive emotion). After the Space-boy offers the Angel a trip to the moon, her emotional state comprises two active emotions, *joy* (intensity 2) and *attraction* (intensity 3), and she expresses the emotion with the higher intensity (see Fig. 3). As we set the decay rate to a high level, all previously elicited emotions (including *distress* and *aversion*) are not part of the Angel's emotional state.

Although we believe that positive attitude and close social distance should have on the agent's emotion expression, it is not clear to us, how emotions such as *affection* or *aversion* should be instantiated by actual behavior. Currently, we use a direct way by simply letting the agent declare those emotions, e.g., *affection* as "I enjoy being with you". However, in some cases, attitude/familiarity based emotions might be used as biasing mechanisms for calculating the intensity of emotion expression, rather than emotions that are externalized by behavior.

5 Conclusion

In this paper, we try to animate characters by providing an explicit model of attitude and familiarity change. This effort toward believable and life-like characters complements research on characters' ability to express, recognize, and reason about emotion. Social interactions usually develop in some way, and we believe that change in attitude and social distance of interlocutors are key aspects of this evolution.

The most needed next step is experimentation with the proposed model. We hypothesize that (animated) characters adjusting their affective behavior as a result of the interaction history are more believable (and less boring) than characters that only consider the current interaction context. However, our model of familiarity change will

have to be refined by considering the nature of the topics dealt with in a conversation. The interaction scenarios described in this paper do not allow to account for the topic parameter, as both are highly restricted and task-specific.

In the future, we hope to integrate our model to pet-type characters and web-based interactive characters that maintain believable long-term relationships with users.

Acknowledgements

We would like to express our thanks to the reviewers for their benevolent and motivating comments. This research is supported by the Research Grant (1999-2003) for the Future Program ("Mirai Kaitaku") from the Japanese Society for the Promotion of Science (JSPS).

References

N. H. Anderson. Primacy effects in impression formation using a generalized order effect paradigm. *Journal of Personality and Social Psychology*, 2:1–9, 1965.

Elisabeth André, Martin Klesen, Patrick Gebhard, Steve Allen, and Thomas Rist. Integrating models of personality and emotions into lifelike characters. In *Proceedings International Workshop on Affect in Interactions - Towards a New Generation of Interfaces*, pages 136–149, 1999.

Cynthia Breazeal-Ferrell and Juan Velásquez. Toward teaching a robot "infant" using emotive communication acts. In *Proceedings of Workshop on Socially Situated Intelligence, in conjunction with SAB-98*, 1998.

Penelope Brown and Stephen C. Levinson. *Politeness. Some Universals in Language Usage*. Cambridge University Press, 1987.

Lola Canamero. Building emotional artifacts in social worlds: Challenges and perspectives. In *Proceedings of 2001 AAAI Fall Symposium in Emotional and Intelligent II: The Tangled Knot of Social Cognition*, pages 22–30. AAAI Press, Technical Report FS-01-02, 2001.

Lola Canamero and Jacob Fredslund. I show you how I like you—Can you read my face? *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans*, 31(5):454–459, 2001.

Justine Cassell and Timothy Bickmore. Negotiated collusion: Modeling social language and its relationship effects in intelligent agents, 2001. User Modeling and Adaptive Interfaces. In Press.

Kerstin Dautenhahn. The art of designing socially intelligent agents – science, fiction, and the human in the loop. *Applied Artificial Intelligence, Special Issue on Socially Intelligent Agents*, 12(7–8):573–617, 1998.

Sylvain Descamps, Helmut Prendinger, and Mitsuru Ishizuka. A multimodal presentation mark-up language for enhanced affective presentation. In *Proceedings International Conference on Intelligent Multimedia and Distant Education (ICIMADE-01), Advances in Educational Technologies: Multimedia, WWW and Distance Education*, pages 9–16, 2001.

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

Barbara Hayes-Roth and Patrick Doyle. Animate characters. *Autonomous Agents and Multi-Agent Systems*, 1 (2):195–230, 1998.

Helen McBreen, Paul Shade, Mervyn Jack, and Peter Wyard. Experimental assessment of the effectiveness of synthetic personae for multi-modal e-retail applications. In *Proceedings 4th International Conference on Autonomous Agents (Agents'2000)*, pages 39–45, 2000.

Microsoft. *Developing for Microsoft Agent*. Microsoft Press, 1998.

Andrew Ortony. Value and emotion. In W. Kessen, A. Ortony, and F. Craik, editors, *Memories, thoughts, and emotions: Essays in the honor of George Mandler*, pages 337–353. Hillsdale, NJ: Erlbaum, 1991.

Andrew Ortony, Gerald L. Clore, and Allan Collins. *The Cognitive Structure of Emotions*. Cambridge University Press, 1988.

David Pautler and Alex Quilici. A computational model of social perlocutions. In *Proceedings of COLING-ACL-98*, 1998.

Catherine Pelachaud and Isabella Poggi. Multimodal embodied agents. In *Proceedings Workshop on Multimodal Communication and Context in Embodied Agents, in conj. with Agents-01*, pages 95–100, 2001.

Helmut Prendinger and Mitsuru Ishizuka. Social role awareness in animated agents. In *Proceedings 5th International Conference on Autonomous Agents (Agents-01)*, pages 270–277, 2001.

Helmut Prendinger and Mitsuru Ishizuka. SCREAM: Scripting Emotion-based Agent Minds. In *Proceedings 1st International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS-02)*, 2002. To appear.

Wolfgang Wahlster, Norbert Reithinger, and Anselm Blocher. SmartKom: Multimodal communication with a life-like character. In *Proceedings of 7th European Conference on Speech Communication and Technology (Eurospeech-01)*, pages 1547–1550, 2001.