Speed Dating with an Affective Virtual Agent - Developing a Testbed for Emotion Models

Matthijs Pontier[1, 2], Ghazanfar Siddiqui[1, 2, 3] and Johan F. Hoorn[1]

[1] VU University Amsterdam, Center for Advanced Media Research Amsterdam, De Boelelaan 1081, 1081HV Amsterdam, The Netherlands
[2] VU University Amsterdam, Department of Artificial Intelligence, De Boelelaan 1083, 1081HV Amsterdam, The Netherlands
[3]Quaid-i-Azam University Islamabad, 45320, Pakistan
{mpr210, ghazanfa}@few.vu.nl
http://www.few.vu.nl/~{mpr210, ghazanfa}

Abstract. In earlier studies, user involvement with an embodied software agent and willingness to use that agent were partially determined by the aesthetics of the design and the moral fiber of the character. We used these empirical results to model agents that in their turn would build up affect for their users much the same way as humans do for agents. Through simulations, we tested these models for internal consistency and were successful in establishing the relationships among the factors as suggested by the earlier user studies. This paper reports on the first confrontation of our agent system with real users to check whether users recognize that our agents function in similar ways as humans do. Through a structured questionnaire, users informed us whether our agents evaluated the user's aesthetics and moral stance while building up a level of involvement with the user and a degree of willingness to interact with the user again.

Keywords: Cognitive Modeling, Emotion Modeling, Speed Dating, Virtual Humans, Empirical Testing

1 Introduction

In prior work, we described how certain dimensions of synthetic character design were perceived by users and how they responded to them [20]. A series of user studies resulted into an empirically validated framework for the study of user-agent interaction with a special focus on the explanation of user engagement and use intentions. We put together results of various studies so as to clarify the individual contributions of agent's affordances, ethics, aesthetics, facial similarity, and realism to the use intentions and engagement of the human user. The interactions between these agent features and their contribution to human user's perception of agents were summarized in a schema called Interactively Perceiving and Experiencing Fictional Characters (I-PEFiC). To date, this framework has a heuristic value because the extracted guidelines are important for anyone who designs virtual characters.

The evidence-based guidelines of I-PEFiC strengthen (e.g., 'beauty leads to involvement') and demystify ('ethics is more important than realism') certain views regarding synthetic character design, giving them experimental support. For example, not only a system's interaction possibilities have a direct effect on the willingness to use an agent but the ethical behavior of the agent as a personality does so too. Moreover, moral behaviors of the agent will be checked for relevance to user's goals and concerns and through that side-track, exert indirect effects on use intentions as well.

In a simulation study [12], we were capable of formalizing the I-PEFiC framework and make it the basic mechanism of how agents and robots build up affect for their human users. In addition, we designed a special module for affective decision making (ADM) which made it possible for the agent to select actions for or against its user.

When we compared I-PEFiC^{ADM} to EMA [8], [15] and CoMERG [4], we found out that the models are complementary to each other. For instance, CoMERG covers a wide variety of emotion regulation strategies. EMA, on the other hand, contains very sophisticated mechanisms for both appraisal and coping, and generating specific emotions based on this appraisal. Therefore, we concluded it would make sense to integrate them [3].

We integrated the three models into Silicon Coppélia [17], and performed simulation experiments to test the behavior of the model. The robotic behavior based on Silicon Coppélia was consistent with the theories the model is based on, and seemed compelling intuitively. However, we tested our models using agent's interacting with each other, not with a real user.

In order to be able to do so, we developed a speed dating application as a testbed for emotion models. In this application, the user interacts with a virtual human on a website. We chose the emotionally laden setting of the speed date because that would make it easy to ask the user what the invisible counterpart would think of them, ethically, aesthetically, and whether they believed the other would want to make an appointment with them, etc.

This testbed served for the first confrontation of Silicon Coppélia with actual users. To make it fit the speed dating domain, we made some changes to Silicon Coppélia, which are described in section 3. We implemented this changed model in the virtual human, thereby enabling it to behave emotionally human-like.

We focused on five factors of Silicon Coppélia that are particularly of interest to a speed-date situation: Good looks (factor Aesthetics), moral behavior (factor Ethics), relevance to personal concerns (Relevance) [8], feeling involved (Involvement), and willingness to meet again (Use Intentions). We wondered whether users would recognize that agents were making ethical and aesthetic assessments and that these assessments were affecting the agent's level of involvement with them as a dating partner as well as the agent's intentions to use (i.e., meet) them again either in another dating session or in real life. Therefore, with our speed-dating application in which the agent was represented by an embodied avatar, we wished to test the following hypotheses.

H1: Users recognize a direct positive effect of agent-assessed Aesthetics on the agent's Involvement with the user

H2: Agent-assessed Ethics of the user has a positive direct effect on Use Intentions of the agent to meet the user again

H3: Relevance of user behavior to agent concerns has a mediating effect on the relation between agent-assessed Ethics and Use Intentions (see H2)

2 Models incorporated in the agent

As suggested in [3], three models were integrated into Silicon Coppélia: EMA [8], [15], CoMERG [4], and I-PEFiC^{ADM} [12]. For this study, we implemented Silicon Coppélia into a speed dating agent. Some changes were made to the model to make it fit the speed dating domain. This section will shortly describe Silicon Coppélia, thereby focusing on the parts where changes were made. More detailed descriptions of Silicon Coppélia and the models it is based on can be found in [3], [4], [8], [12], [15], [17]. Figure 1 shows Silicon Coppélia in a graphical format.

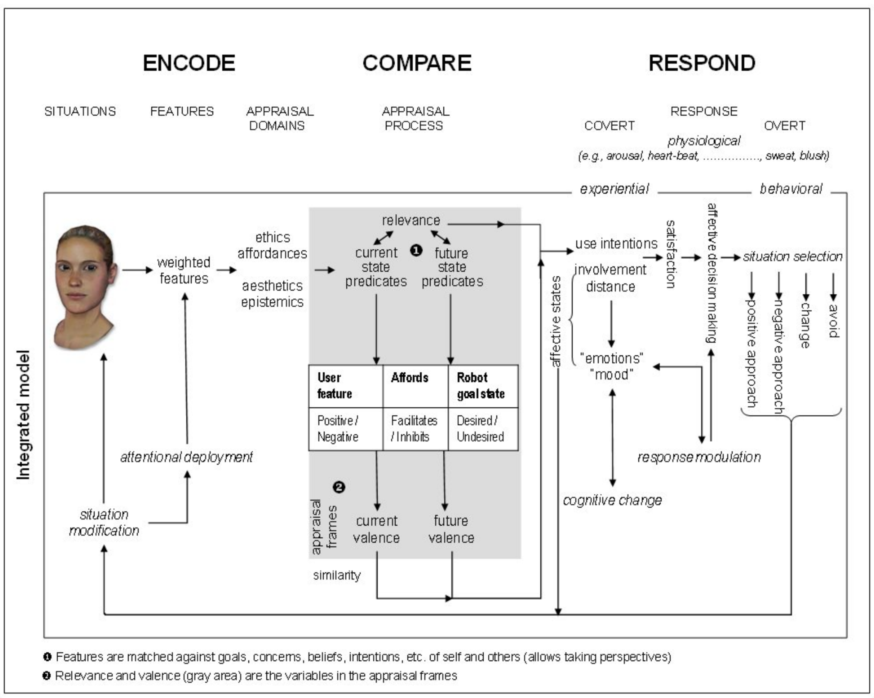

Fig. 1. Integration of CoMERG, EMA and I-PEFiC^{ADM} in “Silicon Coppélia”.

Silicon Coppélia consists of a loop with a situation as input, and actions as output, leading to a new situation. In this loop there are three phases: the encoding, the comparison and the response phase.

In the *encoding* phase, the agent receives others in terms of ethics (good vs. bad), affordances (aid vs. obstacle), aesthetics (beautiful vs. ugly), and epistemics (realistic vs. unrealistic). The agent can be biased in this perception process. The agent has desires with a certain strength for reaching or preventing goal-states that are defined in the system.

In the *comparison* phase, the agent uses beliefs that actions facilitate or inhibit these desired or undesired goal-states to calculate a general expected utility of each action. Further, the perceived features in others, and certain appraisal variables, such as the belief that someone is responsible for reaching or not reaching a goal-state, are appraised for relevance (relevant or irrelevant) and valence to the agent's goals (positive or negative outcome expectancies).

In the *response* phase of the model, this leads to feelings of involvement and distance towards the other, and to use intentions: the agent's willingness to employ the other as a tool to achieve its own goals. Note that with response both overt (behavioral) and covert (experiential) responses are meant in this phase. Emotions such as hope, joy and distress are generated using appraisal variables like the perceived likelihood of goal-states. The agent uses an affective decision making module to calculate the expected satisfaction of possible actions. In this module, affective influences and rational influences are combined in the decision making process. Involvement and distance represent the affective influences, while use intentions and general expected utility represent the more rational influences. When the agent picks and performs an action, a new situation emerges, and the model starts at the first phase again.

3 The Speed-Date Application

We designed a speed date application in which users could interact with a virtual human, named Tom, to get acquainted and make an appointment. The dating partner was performed by our software agent based on Silicon Coppélia, and represented by an avatar created in Haptek's PeoplePutty software [11].

During the speed date, partners could converse about seven topics: (1) Family, (2) Sports, (3) Appearance, (4) Hobbies, (5) Music, (6) Food, and (7) Relationships. For each topic, the dating partners went through an interaction tree with responses that they could select from a dropdown box. To give an idea of what the interaction trees look like, we put the tree for the topic relationships online as an appendix [21].

The agent is capable of simulating five emotions: hope, fear, joy, distress, and anger, which were expressed through the face of the avatar with either a low or a high intensity (depending on little or much relevance of user choices to agent's goals and concerns). Like this, we created 32 (2^5) different emotional states in PeoplePutty; one for each possible combination of two levels of intensity of the five simulated emotions.

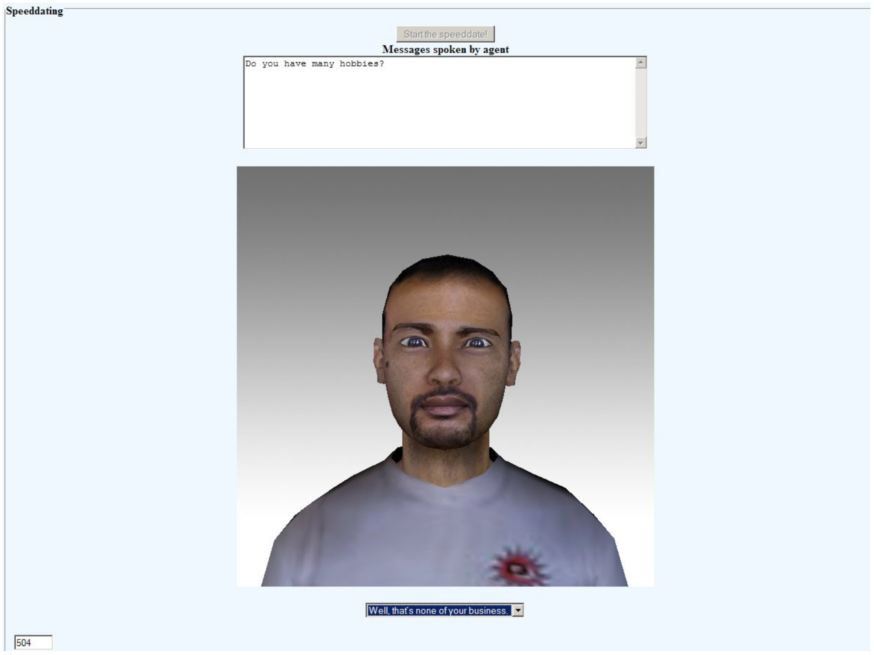

Fig. 2. The speed-dating application.

We created a Web page for the application (Figure 2), in which the virtual agent was embedded as a Haptek player. We used JavaScript [1], a scripting language, in combination with scripting commands provided by the Haptek software [11], to control the Haptek player within the Web browser. In the middle of the Web site, the affective conversational agent was shown, communicating messages through a voice synthesizer and additionally shown in a text area right above the avatar. Figure 2 shows that the avatar looks mildly angry in response to the user's reply "Well, that's none of your business". When the 'start speed date' button above the text area was pressed, the agent introduced itself and started by asking the user a question. The user selected an answer from the dropdown box below the agent. Then the agent responded and so on until the interaction-tree was traversed. When a topic was done, the user could select a new topic or let the agent pick one. When all topics were completed, the message "the speed dating session is over" was displayed and the user was asked to fill out the questionnaire.

3.1 Determining which responses to select

In the speed dating application, the agent perceives the user according to I-PEFiC^{ADM} [12]. It had beliefs that features of the user influenced certain goal-states in the world. For our speed-date setting, the possible goal-states are 'get a date', 'be honest', and 'connecting well' on each of the conversation topics. The agent has beliefs in the range [-1, 1] about the facilitation of these goal-states by each possible response. Further, the agent attaches a general level of positivity and negativity to each response.

During the speed date, the agent updates the assessed goodness of the human user according to the positivity it perceives in the user's responses. It updates the agent-assessed badness on the basis of the negativity it perceives in the user's responses. The ethical values are updated according to the following formulae:

$$\text{New\_Perc(Good)} = \beta_{eth} * \text{Old\_Perc(Good)} + (1-\beta_{eth}) * \text{Perc(Positivity\_Response)}$$

$$\text{New\_Perc(Bad)} = \beta_{eth} * \text{Old\_Perc(Bad)} + (1-\beta_{eth}) * \text{Perc(Negativity\_Response)}$$

In these formulae, the persistency factor β_{eth} is the proportion of the old perceptions that is taken into account to determine the new perceptions. The remaining part of the perceptions is determined by the positivity and negativity in the response of the human user.

The agent establishes beliefs about the beauty of the human user in the domain [0, 1], based on user responses during the conversation topic 'appearance'. The agent updates the assessed beauty and ugliness of the user, using the following formulae. In these formulae, β_{aesth} serves as the persistency factor, and Perc(Beauty_Response) represents the perceived beauty of the user based on the user's response.

$$\text{New\_Perc(Beauty)} = \beta_{aesth} * \text{Old\_Perc(Beauty)} + (1-\beta_{aesth}) * \text{Perc(Beauty\_Response)}$$

$$\text{New\_Perc(Ugly)} = \beta_{aesth} * \text{Old\_Perc(Ugly)} + (1-\beta_{aesth}) * (1-\text{Perc(Beauty\_Response)})$$

I-PEFiC^{ADM} has an affective decision module in which rational and affective dimensions are combined to make an affective decision. To determine the rational dimensions, the general expected utility (GEU) [8] with respect to the goals is calculated. The agent also calculates the GEU of the actions of the human participant regarding the agent's own goals. The agent uses this to determine the perceived affordances of the participant in terms of being an aid or an obstacle in achieving the agent's goals, using the following formulae:

IF GEU_action(Human_Action) > 0

THEN

$$\text{New\_Perc(Aid)} = \text{Old\_Perc(Aid)} + \alpha_{aff} * \text{GEU(Human\_Action)} * (1-\text{Old\_Perc(Aid)})$$

$$\text{New\_Perc(Obst)} = \text{Old\_Perc(Obst)} - \alpha_{aff} * \text{GEU(Human\_Action)} * \text{Old\_Perc(Obst)}$$

IF GEU_action(Human_Action) < 0

THEN

$$\text{New\_Perc(Aid)} = \text{Old\_Perc(Aid)} + \alpha_{aff} * \text{GEU(Human\_Action)} * \text{Old\_Perc(Aid)}$$

$$\text{New\_Perc(Obst)} = \text{Old\_Perc(Obst)} - \alpha_{aff} * \text{GEU(Human\_Action)} * (1-\text{Old\_Perc(Obst)})$$

In these formulae, α_{aff} is a *modification factor* that determines how quickly the variable was updated. This modification factor is multiplied with the *impact value* GEU(Human_Action). Multiplying with a *limiter* (in the case of the first formula this is (1-Old_Perception(Aid))) avoids that the formula goes out of range. It also manages that if an agent's assessments approach an extreme value, it will be harder to push it further to the extreme and easier to move it back to a less extreme value.

Our speed-dating agent is capable of developing a bias [0, 2] in the assessment of the moral fiber of the user. This bias is initiated at a neutral value of 1. In [17], the believed responsibility of oneself and others for reaching certain goal states is calculated. This believed responsibility is used to calculate a bias in perceiving the ethics of the human user, using the formulae presented next. Note that these formulae have the same form (with a modification factor, an impact value, and a limiter) as those for calculating the perceived aid and obstacle described earlier:

IF -Belief(Human_Responsible, Goal) * Ambition(Goal) > 0

THEN New_Bias(Good) = Old_Bias(Good) - $\alpha_{b\text{-}eth}$ * Ambition(Goal) * Old_Bias(Good)

New_Bias(Bad) = Old_Bias(Bad) + $\alpha_{b\text{-}eth}$ * Ambition(Goal) * (2 - Old_Bias(Bad))

IF -Belief(Human_Responsible, Goal) * Ambition(Goal) < 0

THEN New_Bias(Good) = Old_Bias(Good) + $\alpha_{b\text{-}eth}$ * -Ambition(Goal) * Old_Bias(Good)

New_Bias(Bad) = Old_Bias(Bad) - $\alpha_{b\text{-}eth}$ * -Ambition(Goal) * (2 - Old_Bias(Bad))

Using these variables, the agent determines its response. The agent-assessed ethics, aesthetics, realism, and affordances of the user lead, while matching these aspects with the goals of the agent, to 'feelings' of Involvement and Distance towards the human user and a general expected utility of each action, as described in [17]. Each time, the agent can select its response from a number of options. The expected satisfaction of each possible response is calculated based on the Involvement and Distance towards the user and the general expected utility of the response, using the following formula:

$$\begin{aligned}\text{ExpectedSatisfaction}_{(\text{Action})} = \;& w_{eu} * \text{GEU}_{(\text{Action})} + \\ & w_{pos} * (1 - \text{abs}(\text{positivity} - \text{bias}_I * \text{Involvement})) + \\ & w_{neg} * (1 - \text{abs}(\text{negativity} - \text{bias}_D * \text{Distance}))\end{aligned}$$

The agent searches for an action with the level of positivity that comes closest to the level of (biased) involvement, with the level of negativity closest to the level of (biased) distance, and with the strongest action tendency.

3.2 Determining which emotions to express

During the speed date, the agent simulates a series of emotions, based on the responses given by the user. Hope and fear are calculated each time the user gives an answer. Hope and fear of the agent are based on the perceived likelihood that the agent will get a follow-up date. The likelihoods are used in the following function to calculate the hope for achieving a goal. This function is similar to the function described in [5].

IF f >= likelihood

THEN hope_for_goal = -0,25 * (cos(1 / f * π * likelihood(goal)) -1,5) * ambition(goal)

IF f < likelihood

THEN hope_for_goal = -0.25 * (cos(1 / (1-f) * π * (1-likelihood(goal))) -1.5) * ambition(goal)

In these functions, f is a shaping parameter (in the domain [0, 1]) that can be used to manipulate the location of the top of the hope curve. The value of this parameter may differ per individual, and represents 'fatalism' (or 'pessimism'): The top of the likelihood/hope-curve is always situated at the point where likelihood = f. Thus, for an f close to 1, the top of the curve is situated to the extreme right (representing persons that only 'dare' to hope for events that are likely to happen). Similarly, for an f close to 0, the top of the curve is situated to the extreme left (representing persons that already start hoping for very unlikely events). In our current study, f was set at 0.5. We created a smooth function instead of a linear function, because this matches the emotion curves found for humans the best. Further, a higher ambition level leads to higher hopes. If the ambition level is negative (i.e., the goal is undesired), the outcome of hope_for_goal will be a negative value.

The following algorithm is executed on the found values for hope_for_goal:

1. Sort the values in two lists: [0→1] and [0→-1]
2. Start with 0 and take the mean of the value you have and the next value in the list. Continue until the list is finished. Do this for both the negative and the positive list.
3. Hope = Outcome positive list. Fear = abs(Outcome negative list).

The values are sorted in a list with positive hope_for_goal's (i.e., hope for desired goals), and negative hope_for_goal's (i.e., fear for undesired goals). For both the lists, 0 is the starting point and the mean of the value you have and the next value in the list (where the next value is the value closest to 0 that is left in the list) is picked until the end of the list is reached. The new level of hope for the agent is the outcome of the positive list and the new level of fear for the agent is the absolute value of the outcome of the negative list.

The joy and distress of the agent are based on reaching or not reaching desired or undesired goal-states. If a goal-state becomes true (i.e., the agent matches well with the user on a certain conversation topic), the levels of joy and distress are calculated by performing the following formulae:

IF (ambition (goal) * belief(goal)) > 0

THEN new_joy = old_joy + mf_joy * ambition(goal) * belief(goal) * (1-old_joy)

new_distress = old_distress + mf_distress * -ambition(goal) * belief (goal) * old_distress

IF (ambition (goal) * belief (goal)) < 0

THEN new_joy = old_joy + mf_joy * ambition(goal) * belief(goal)* old_joy

new_distress = old_distress + mf_distress * -ambition(goal) * belief (goal) * (1-old_distress)

In these formulae, mf_joy and mf_distress are modification factors that determine how quickly joy and distress are changed if the agent reaches a certain goal-state. In this paper, the values were both set to 1. These modification factors are multiplied with the impact value, which is ambition(goal) for joy and -ambition(goal) for distress. This way, if a desired goal is achieved, joy is increased and distress is decreased. Conversely, achieving an undesired goal decreases joy and increases distress. Multiplying with limiter (1-old_joy) for joy and old_distress for distress if the goal is desired, keeps the formula from going out of range and drives possible extreme values of joy and distress back to a milder level. If the achieved goal-state is undesired, old_joy is used as limiter for joy and (1-old_distress) as a limiter for distress, because the values of joy and distress will move into the opposite direction of when the goal is desired.

The anger of the agent is calculated using the believed responsibility of the human user for the success of the speed date:

IF Belief(Human_Responsible, Goal)) * Ambition(Goal) > 0

THEN Anger(Agent) = old_anger + mf_{anger} * (-Belief(Human_Responsible, Goal)) * Ambition(Goal) * (1 - old_anger)

IF Belief(Human_Responsible, Goal)) * Ambition(Goal) < 0

THEN Anger(Agent) = old_anger + mf_{anger} * (-Belief(Human_Responsible, Goal)) * Ambition(Goal) * old_anger

To calculate the level of the agent's anger with the user in the range [0, 1], the above formula is used, with Belief(Human_Responsible, Goal) * Ambition(Goal) as *impact value*. This way, if an agent believes a desired goal state should have been reached, but it has not, the agent will become angrier with the user who the agent holds responsible for not achieving the desired goal state. The agent will become less angry with the user who is believed to have tried helping to reach the desired goal state. The reverse happens, when the goal state is undesired. Because people do not stay angry forever, anger is multiplied with a decay factor at each time step.

All five emotions implemented into the system (i.e., hope, fear, joy, distress, and anger) are simulated in parallel (see [20]). If the level of joy, distress, or anger is below 0.5, a low intensity of the emotion was facially expressed by the agent. If the level of joy, distress, or surprise was greater or equal than 0.5, a high intensity of the emotion was expressed by the agent. Because within the given parameter settings, hope and fear rarely reached extreme values, this boundary was set to 0.25 for hope and fear.

4 Experiment

To examine the hypotheses H1-3, a user study was performed in which users diagnosed the cognitive mechanisms by which their robotic dating partner came to an assessment of and affective attitude towards its user.

4.1. Participants

A total of 18 participants ranging in age from 18 to 24 years (M = 19.17, SD = 1.86) volunteered for course credits. All participants were Dutch female students of VU University Amsterdam (Communication Science Program). We confronted female participants with male agents, because the ability to describe emotional feelings as clearly as possible was considered a prerequisite to assess the agent's affective performance. Previous research suggests that women are better equipped to do an emotional assessment of others [2]. Participants were asked to rate their experience in dating and computer-mediated communication on a scale from 1 to 7 and appeared to be reasonably experienced in dating (M = 4.00, SD = 1.72) and communicated frequently via a computer (M = 5.50, SD = 1.04).

4.2 Procedure

Preceding the experiment, the participants were told a cover story that they were participating in a speed-dating session with a male participant, named Tom, who was at another location. After a session of about 10 minutes, the participants filled out a questionnaire of 94 items. At the end of the experiment, the participants were thanked for their participation and were debriefed that the avatar they communicated with represented not a real human but a computer-generated agent used to test our affect generation and regulation software.

4.3 Measurements

We developed a 94-item questionnaire in Dutch in which the dependent variables of H1-3 were incorporated. All items were in the form of statements of the type "I think that Tom finds me attractive" measured on a 7-point Likert-type rating scale, which ranged from 'totally disagree' (1) to 'totally agree' (7). Negatively phrased items were reverse-coded before entering the analysis.

Ethics measured in how far the participants perceived the agent as good or bad, using the four items 'trustworthy', 'credible', 'malicious' and 'mean'. The scale appeared to be reliable (Cronbach's α = .83). *Aesthetics* measured in how far the participant perceived the agent as beautiful or ugly, using the four items 'attractive', 'good-looking', 'ugly' and 'bad-looking'. The scale appeared to be very reliable (Cronbach's α = .94). *Relevance* was measured by three items indicating how important or useless the user was in 'creating a good atmosphere' and in 'completing each other during conversation'. This scale appeared to be very reliable (Cronbach's α = .88). *Use Intentions* were measured with four items. These were 'happy to have met', 'wanting to meet in another context', 'sad to have met' and 'wanting to get rid of'. The scale appeared to be very reliable (Cronbach's α = .90). *Involvement* was measured with four items, which were 'appeal', 'good feeling', 'feeling connected' and 'feeling engaged'. Also this scale appeared to be very reliable (Cronbach's α = .93).

The remaining items referred to sociodemographic variables or were derived from questionnaires dedicated to other emotion regulation models [10], [19], [20] so to assess the psychometric quality of the items, which was on the whole quite disappointing. The results of the pre-test of these additional items fall beyond the scope of the present paper and will be reported elsewhere.

5 Analysis and Results

The items on each scale were averaged for each participant and the grand mean scale values across participants were used in a series of linear regressions to explore the hypotheses H1-3.

H1 predicted that users would see that the agent assessed their looks (factor Aesthetics), which was then used by the agent to determine a level of Involvement with the user. The direct relation between agent-assessed Aesthetics (M = 4.10, SD = 1.32) and Involvement (M = 2.46, SD = 1.24) was analyzed using linear regression. H1 was confirmed in that Aesthetics predicted Involvement ($R^2 = .708$, $r(16) = .85$, $p < .01$) indicating a significant positive relation between the two variables. Therefore, H1 could be confirmed.

H2 and H3 were tested in unison, stating that there should be direct effects of agent-assessed Ethics of the user on Use Intentions (H2) complemented by indirect effects of Ethics through the Relevance of the user to agent's goals and concerns (H3). We performed a Sobel-test [18] to investigate the effects of Ethics (M = 4.10, SD = .92) on Use Intentions (M = 3.65, SD = 1.24) and the predicted mediating role of Relevance (M = 4.08, SD =.92). Ethics served as the independent variable, Relevance as the mediator, and Use Intentions as the dependent variable. The results showed a significant direct effect of Ethics on Use Intentions (Sobel $z = .88$, $t(16) = 2.51$, $p < .05$), supporting H2. However, no significant direct effects were found between Ethics and Relevance or Relevance and Use Intentions. The predicted indirect effect of Ethics through Relevance was not significant either (Sobel $z = 28$, $p = .24$), so that H3 was rejected.

6 Discussion

We developed a speed-dating application, in which a human user could communicate with an affective agent via multiple-choice questions on a website. Because of the emotionally laden setting of a speed date, this application serves well as a testbed for emotion models. Many emotion models work with (interrelated) goals. In this application, already some goals and their relation has been pre-defined. Further, using the Haptek software [11], the agent can easily be modified, for example by changing the appearance, the voice, and the types and intensities of facial expressions shown by the agent.

In a previous study, various emotion models, such as [6], [16], were validated by comparing human self-reported emotion intensities while playing a game with predictions of the emotion models [9]. However, this did not involve letting humans judge the behavior generated by the models as a validation method.

In the current study, we used the application to test Silicon Coppélia [17], an affect generation and regulation system that builds up an affective attitude of the agent towards its user, and makes decisions based on rational as well affective influences. In a user study with young females, we assessed in how far users recognized that a male agent, named Tom, was responding according to the predictions by Silicon Coppélia, described in the introduction.

H1 was confirmed in that female users recognized a direct positive effect of Tom's assessed aesthetics of the female user on Tom's involvement with her. Put differently, females recognized that when Tom found them attractive, he became more involved in terms of having 'good feelings' about them and 'feeling connected'. Note that Tom did not explicitly express these affective states but that the female users were under the impression that this is how Tom felt about them. By confirming H1, we found the mirror-image of a well-established empirical finding [20] that users become more involved with an embodied agent when they find it aesthetically more pleasing. We now have confirmation that on this aspect, our software behaves in a human-like fashion.

H2 was confirmed in that the female users recognized that Tom assessed their moral fiber and that this had a direct and positive effect on Tom's intentions to use (or better meet) the user again, either in another dating session or offline. In other words, female users saw that Tom was more inclined to meet them again when he thought they were more 'credible' and 'trustworthy'. Again, the confirmation of H2 is the mirror image of earlier findings [20], indicating that our software assesses humans like humans assess embodied software agents.

H3 was refuted in the sense that ethical aspects of the female user were not played through relevance to Tom's goals in order to evoke effects on Tom's use intentions. Female users only saw the direct effects of Tom's ethical assessment on how he felt about them (H2). The absence of the mediating effect of relevance shows the limitations of our software in authentically simulating human-like affective behavior. Mimicking indirect affective mechanisms seem to be a bridge too far. For now.

There are two problems to crack if we want to explain the absence of the effect of relevance as a mediating variable. First, it might be that the items were not indicating relevance as a concept well enough, although they were psychometrically congruent. However, this might mean that although the items measured the same, it was not relevance to the agent's goals and concerns that they measured. Thus, better concept analysis and pre-testing of items should improve the contents of the relevance scale. Second, it might be that the behavioral representation of relevance in the interaction options was not expressed well enough. This would take a round of redesign and user testing to see whether there is a way to surpass a certain threshold of recognition.

Future research with human users, however, should not only focus on ethics, aesthetics, involvement and use intentions. Although important, they are flanked by other factors such as realism and affordances [20]. Further, it is not yet assessed whether models such as EMA [8], [14], [15] or Gross [4] have added value for producing affective behavior that is recognizable for humans.

The application is generic in the sense that (emotion) models can easily be connected to the agent. After connecting, the agent bases its behavior (i.e., speech acts, facial expressions, etc.) on these models. Thereby, if the models are correct, it should be able to simulate behavior that is perceived as emotionally human-like by humans. By implementing several (versions of) models into the agent, multiple test-conditions can be created to compare the different models.

Also, the application can easily be adjusted to let a human control the speed dating agent, which enables doing Wizard of Oz studies [13]. In future research, we plan to compare the performance of models such as I-PEFiC^{ADM}, EMA, and Gross with a Wizard of Oz condition of human-human interaction, which allows for making stronger claims to the behavioral fidelity of an agent's affective response mechanisms.

Acknowledgements

We would like to thank Tibor Bosse for commenting on earlier drafts of this paper. We are also grateful to the master students H. Blok, J. Jobse, S. Bor, Y. Kobayashi, R. Changoer, D. Langeveld, P. Dirks, B. Rond, K. van Gool, and E. Stam for their help in designing the interactions between the agent and its user, for their help in developing and testing the questionnaire, and for performing statistical analyses.

References

1. About JavaScript – MDC, http://developer.mozilla.org/en/docs/About_JavaScript
2. Barrett, L.F., Robin, L., Pietromonaco, P.R., Eyssell, K.M.: Are Women The "More Emotional Sex?" Evidence from Emotional Experiences in Social Context. In: Cognition and Emotion, vol. 12, pp. 555--578 (1998)
3. Bosse, T., Gratch, J., Hoorn, J.F., Pontier, M.A., Siddiqui, G.F.: Comparing Three Computational Models of Affect In: Proceedings of the 8th International Conference on Practical Applications of Agents and Multi-Agent Systems , PAAMS, pp. 175--184 (2010)
4. Bosse, T., Pontier, M.A., Treur, J.: A Dynamical System Modelling Approach to Gross´ Model of Emotion Regulation. In: Lewis, R.L., Polk, T.A., and Laird, J.E. (eds.), Proceedings of the 8th International Conference on Cognitive Modeling, ICCM'07, pp. 187--192, Taylor and Francis (2007)
5. Bosse, T., Zwanenburg, E.: There's Always Hope: Enhancing Agent Believability through Expectation-Based Emotions. In: Pantic, M., Nijholt, A., Cohn, J. (eds.), Proceedings of the 2009 International Conference on Affective Computing and Intelligent Interaction, ACII'09, pp. 111--118, IEEE Computer Society Press (2009)
6. Dias, J. and Paiva, A.: Feeling and Reasoning: a Computational Model for Emotional Agents. In: 12th Portuguese Conference on Artificial Intelligence, Springer (2005)
7. Frijda, N.H., Swagerman, J.: Can Computers Feel? Theory and Design of an Emotional System. In: Cognition and Emotion, vol. 1, pp. 235--258 (1987)
8. Gratch, J., Marsella, S.: Evaluating a computational model of emotion. In: Journal of Autonomous Agents and Multiagent Systems (Special issue on the best of AAMAS 2004), vol. 11, no. 1, pp. 23--43 (2006)
9. Gratch, J., Marsella, S., Wang, N., Stankovic, B.: Assessing the validity of appraisal-based models of emotion. In: Pantic, M., Nijholt, A., Cohn, J. (eds.), Proceedings of the International Conference on Affective Computing and Intelligent Interaction, ACII'09, IEEE Computer Society Press (2009)

10. Gross, J.J.: Emotion Regulation in Adulthood: Timing is Everything. In: Current Directions in Psychological Science, 10(6), pp. 214--219 (2001)
11. Haptek, Inc., http://www.haptek.com
12. Hoorn, J.F., Pontier, M.A., Siddiqui, G.F.: When the user is instrumental to robot goals. First try: Agent uses agent. In: Proceedings of IEEE/WIC/ACM Web Intelligence and Intelligent Agent Technology 2008, WI-IAT '08, IEEE/WIC/ACM, Sydney AU, pp. 296--301 (2008)
13. Landauer, T.K.: Psychology as a mother of invention. In: Proceedings of the SIGCHI/GI conference on Human factors in computing systems and graphics interface, pp. 333--335 (1987)
14. Marsella, M., Gratch, J.: EMA: A Process Model of Appraisal Dynamics. In: Journal of Cognitive Systems Research, Vol. 10, No. 1, pp. 70--90 (2009)
15. Marsella, S., Gratch, J.: EMA: A Model of Emotional Dynamics. In: Cognitive Systems Research, vol. 10, no. 1, pp. 70--90 (2009)
16. Neal Reilly, W.S.: Believable Social and Emotional Agents.CMU: Pittsburgh, PA. (1996)
17. Pontier, M.A., Siddiqui, G.F.: Silicon Coppélia: Integrating Three Affect-Related Models for Establishing Richer Agent Interaction. In: wi-iat, IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, vol. 2, pp. 279--284, (2009)
18. Preacher, K. J., and Hayes, A.F.: SPSS and SAS procedures for estimating indirect effects in simple mediation models. In: Behavior Research Methods, Instruments, & Computers, 36 (4), pp. 717--731 (2004)
19. Smith, C.A., Lazarus, R.S.: Emotion and Adaptation. In: L.A. Pervin (Ed.), Handbook of Personality: Theory & Research. NY, USA Guilford Press, pp. 609--637 (1990)
20. Van Vugt, H.C., Hoorn, J.F., Konijn, E.A.: Interactive engagement with embodied agents: An empirically validated framework. In: Computer Animation and Virtual Worlds, Vol. 20, pp. 195--204 (2009)
21. http://www.few.vu.nl/~ghazanfa/IVA2010/