How Women Think Robots Perceive Them – as if Robots were Men

Matthijs A. Pontier and Johan F. Hoorn

Center for Advanced Media Research Amsterdam (CAMeRA@VU) / The Network Institute, VU University,
De Boelelaan 1081, 1081HV, Amsterdam, The Netherlands,

Abstract: In previous studies, we developed an empirical account of user engagement with software agents. We formalized this model, tested it for internal consistency, and implemented it into a series of software agents to have them build up an affective relationship with their users. In addition, we equipped the agents with a module for affective decision-making, as well as the capability to generate a series of emotions (e.g., joy and anger). As follow-up of a successful pilot study with real users, the current paper employs a non-naïve version of a Turing Test to compare an agent's affective performance with that of a human. We compared the performance of an agent equipped with our cognitive model to the performance of a human that controlled the agent in a Wizard of Oz condition during a speed-dating experiment in which participants were told they were dealing with a robot in both conditions. Participants did not detect any differences between the two conditions in the emotions the agent experienced and in the way he supposedly perceived the participants. As is, our model can be used for designing believable virtual agents or humanoid robots on the surface level of emotion expression.

1 INTRODUCTION

1.1 Background

There is a growing interest in developing embodied agents and robots. They can make games more interesting, accommodate those who are lonely, provide health advice, make online instructions livelier, and can be useful for coaching, counselling, and self-help therapy. In extreme circumstances, robots can also be the better self of human operators in executing dangerous tasks.

For a long time, agents and social robots were mainly developed from a technical point of view but we now know it is not a matter of technology alone. Theories and models of human life are also important to explain communication rules, social interaction and perception, or the appraisal of certain social situations. In media psychology, mediated interpersonal communication and human-computer interaction, emotions play a salient role and cover an important area of research (Konijn and Van Vugt, 2008).

The idea of affective computing (Picard, 1997) is that computers 'have' emotions, and detect and understand user emotions to respond appropriately to the user. Virtual agents who show emotions may increase the user's likeability of a system. The positive effects of showing empathetic emotions are repeatedly demonstrated in human-human communication (e.g., Konijn and Van Vugt, 2008) and are even seen as one of the functions of emotional display. Such positive effects may also hold when communicating with a virtual agent. Users may feel emotionally attached to virtual agents who portray emotions, and interacting with such "emotional" embodied computer systems may positively influence their perceptions of humanness, trustworthiness, and believability. User frustration may be reduced if computers consider the user's emotions (Konijn and Van Vugt, 2008). A study by Brave et al. (2005) showed that virtual agents in a blackjack computer game who showed empathic emotions were rated more positively, received greater likeability and trustworthiness, and were perceived with greater caring and support capabilities than virtual agents not showing empathy.

Compared to human affective complexity, contemporary affective behavior of software agents and robots is still quite simple. In anticipation of emotionally more productive interactions between user and agent, we looked at various models of human affect-generation and affect-regulation, to see how affective agent behavior can be improved.

1.2 From Theories to Computation

Previous work described how certain dimensions of synthetic character design were perceived by users and how they responded to them (Van Vugt et al., 2009). A series of user studies into human-agent interaction resulted into an empirically validated framework called Interactively Perceiving and Experiencing Fictional Characters (I-PEFiC). I-PEFiC explains the individual contributions and the interactions of an agent's Affordances, Ethics, Aesthetics, facial Similarity, and Realism to the Use Intentions and Engagement of the human user. To date, this framework has an explanatory as well as a heuristic value because the extracted guidelines are important for anyone who designs virtual characters.

In a simulation study (Hoorn et al., 2008), 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) that made it possible for the agent to select actions in favor or against its user, hence I-PEFiC^{ADM}.

To advance I-PEFiC^{ADM} in the area of emotion regulation, we also looked at other models of affect (Bosse et al., 2010). Gratch and Marsella (2009) formalized the theory of Emotion and Adaptation of Smith and Lazarus (1990) into EMA, to create agents that cope with negative affect. The emotion-regulation theory of Gross (2001) inspired Bosse et al., (2007) to develop CoMERG (the Cognitive Model for Emotion Regulation based on Gross). Together, these approaches cover a large part of appraisal-based emotion theory (Frijda et al.,) and all three boil down to appraisal models of emotion. We therefore decided to integrate these three affect

models into a model we called Silicon Coppélia (Pontier and Siddiqui, 2009; Hoorn et al., 2012). Figure 1 shows Silicon Coppélia in a graphical format.

Silicon Coppélia consists of a loop with a situation as input, and actions as output, leading to a new situation. This loop consists of three phases: (1) encoding, (2) comparison, and (3) response.

In the *encoding* phase, the agent perceives other agents (whether human or synthetic) 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, because it is equipped with desires that have a certain strength for achieving or preventing pre-defined goal-states ('get a date', 'be honest' and 'connect well').

In the *comparison* phase, the agent retrieves beliefs about actions facilitating or inhibiting the desired or undesired goal-states to calculate a general expected utility of each action. Further, agent uses certain appraisal variables, such as the belief that someone is responsible for accomplishing goal-states or not. These variables and the perceived features of others are appraised for Relevance (relevant or irrelevant) and Valence to the agent's goals and concerns (positive or negative outcome expectancies).

In the *response* phase of the model, the resulting appraisals lead to processes of Involvement and Distance towards the other, and to the emergence of certain Use Intentions: The agent's willingness to employ the other as a tool to achieve its own goals. Note that both overt (behavioral) and covert (experiential) responses can be executed in this phase. Emotions such as hope, joy, and anger are generated using appraisal variables such as the perceived likelihood of goal-states. The agent uses an affective
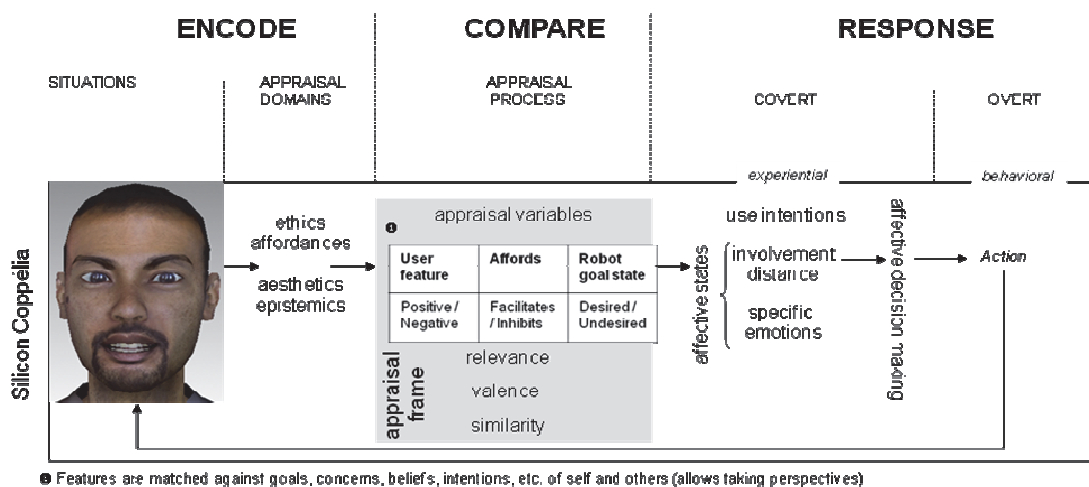

Figure 1: Graphical Representation of Silicon Coppelia.

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, whereas Use Intentions and general expected utility represent the more rational influences. When the agent selects and performs an action, a new situation emerges, and the model starts at the first phase again.

1.3 Speed-dating as a New Turing-test

In previous research we developed a speed-dating application as a testbed for cognitive models (Pontier et al., 2010). In this application, the user interacted with Tom, a virtual agent on a Website.

We opted for a speed-dating application, because we expected this domain to be especially useful for testing emotion models. The emotionally laden setting of the speed-date simplified asking the user what Tom would think of them, ethically, aesthetically, and whether they believed the other would want to see them again, etc. Further, in a speed-date there usually is a relatively limited interaction space; also in our application, where we made use of multiple choice responses. This was done to equalize the difference between a human and our model in the richness of interaction, which was not our research focus. We wanted the difference to be based on the success or failure of our human-like emotion simulations.

We chose to confront female participants with a male agent, because we expected that the limitations in richness of behavior in the experiment would be more easily accepted from a male agent than from a female one. Previous research suggests that men usually have more limited forms of emotional interaction and that women are usually better equipped to do an emotional assessment of others (Barret et al., 1998). By means of a questionnaire, the participants diagnosed the emotional behavior, and the cognitive structure behind that behavior, simulated by our model, or performed by a "puppeteer" controlling Tom.

A pilot study (Pontier et al., 2010) showed that users recognized at least certain forms of human affective behavior in Tom. Via a questionnaire, users diagnosed for us how Tom perceived them and whether they recognized human-like affective mechanisms in Tom. Although Tom did not explicitly talk about it, the participants recognized human-like perception mechanisms in Tom's behavior. This finding was a first indication that our software had a humanoid way of assessing humans,

not merely other software agents.

These results made us conduct a follow-up 'Wizard of Oz' (Landauer, 1987) experiment with 54 participants. In this experiment we compared the performance of Tom equipped with Silicon Coppélia to the performance of a human controlling Tom as a puppeteer. This experiment may count as an advanced version of a Turing Test (Turing, 1950).

In a Turing Test, however, participants are routinely asked whether they think the interaction partner is a human or a robot. In this experiment, however, we did not ask them so directly. After all, because of the limited interaction possibilities of a computer interface, the behavior of Tom may not seem very human-like. Therefore, all participants would probably have thought Tom was a robot, and not a human, making it impossible to measure any differences. Therefore, we introduced the speed-dating partner as a robot to see whether humans would recognize human affective structures equally well in the software and in the puppeteer condition.

Further, when testing the effect of a virtual interaction partner on humans, participants are usually asked how they experience the character. In this experiment, however, we asked people how they thought the character perceived them. Thus, the participants served as a diagnostic instrument to assess the emotional behavior of Tom, and to detect for us the cognitive structure behind that behavior. This way, we could check the differences between our model and a human in producing emotional behavior, and the cognitive structure responsible for that behavior.

We hypothesized that we would not find any differences between the behavior of Tom controlled by our model and that of Tom controlled by a human, indicating the success of Silicon Coppélia as a humanoid model of affect generation and regulation. This would also indicate the aptness of the theories the model is based on. Because Silicon Coppélia is computational, this would also be very interesting for designing applications in which humans interact with computer agents or robots.

2 METHOD

2.1 Participants

A total of 54 Dutch female heterosexual students ranging from 18-26 years of age (M=20.07, SD=1.88) volunteered for course credits or money (5 Euros). Participants were asked to rate their experience in dating and computer-mediated communication on a scale from 0 to 5. Participants

communicated frequently via a computer (M = 4.02, SD = 1.00) but appeared to have little experience in online dating (M = .33, SD = .80).

2.2 Materials: Speed-dating Application

We designed a speed-date application in which users could interact with a virtual agent, named Tom, to get acquainted and make an appointment. The dating partner was represented by Tom, an avatar created in Haptek's PeoplePutty software.

Tom 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. This depended on little or much relevance of user choices to Tom'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.

We created a Web page for the application (see Figure 2), in which the virtual agent was embedded as a Haptek player. We used JavaScript in combination with scripting commands provided by the Haptek software, 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 (e.g., "Do you have many hobbies?") and additionally shown as text right above the avatar. Figure 2 shows that the avatar looks annoyed in response to the user's reply "Well, that's none of your business".
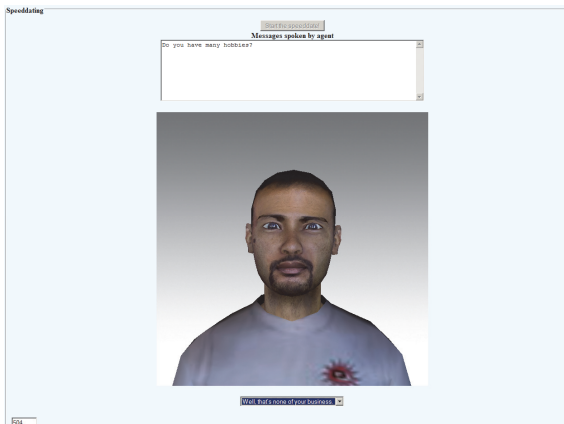

Figure 2: The speed-dating application.

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 inserted the tree for Relationships in the Appendix.

When the 'start speed-date' button above the text area was pressed, Tom introduced himself and started by asking the user a question. The user selected an answer from the dropdown box below Tom. Then Tom responded and so on until the interaction-tree was traversed. When a topic was done, the user could select a new topic or let Tom select 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.

In the speed-dating application, Tom perceived the user according to Silicon Coppélia (Hoorn et al., 2012). Tom had beliefs that features of the user influenced certain goal-states in the world. For our speed-date setting, the possible goal-states were 'get a date', 'be honest', and 'connecting well' on each of the conversation topics. Tom had beliefs about the facilitation of these goal-states by each possible response. Further, Tom attached a general level of positivity and negativity to each response.

During the speed-date, Tom updated its perception of the user based on her responses during the speed-date, as described in (Pontier et al., 2010). The assessed Ethics, Aesthetics, Realism, and Affordances of the user led, while matching these aspects with the goals of Tom, to Involvement and Distance towards the human user and a general expected utility of each action. Each time, Tom selected its response from a number of options. The expected satisfaction of each possible response was calculated based on the Involvement and Distance towards the user and the general expected utility of the response, using the following formula:

$$ExpectedSatisfaction_{(Action)} =$$
$$w_{eu} * GEU_{(Action)} +$$
$$w_{pos} * (1 - abs(positivity - bias_I * Involvement)) +$$
$$w_{neg} * (1 - abs(negativity - bias_D * Distance))$$

Tom searched for an action with the level of positivity that came closest to the level of Involvement, with the level of negativity closest to the level of Distance, and with the highest expected utility (GEU). Tom could be biased to favor positive or negative responses to another agent.

During the speed-date, Tom simulated a series of emotions, based on the responses given by the user. Hope and fear were calculated each time the user gave an answer. Hope and fear of Tom were based on the perceived likelihood that he would get a

follow-up date. The joy and distress of Tom were based on achieving desired or undesired goal-states or not. The anger of Tom was calculated using the assumed responsibility of the human user for the success of the speed-date.

All five emotions implemented into the system (i.e., hope, fear, joy, distress, and anger) were simulated in parallel. If the level of emotion was below a set boundary, a low intensity of the emotion was facially expressed by Tom. If the level of emotion was greater or equal than the boundary, a high intensity of the emotion was expressed by Tom.

2.3 Design

The participants were randomly assigned to two experimental conditions. In the first condition, Tom was controlled by Silicon Coppélia, whereas in the second condition Tom was controlled by a human trained to handle him (Wizard of Oz condition, WOz). All participants assumed they were interacting with a robotic partner; also in the WOz condition. To have some control over the idiosyncrasies of a single human controller, the WOz condition consisted of two identical sub-conditions with a different human puppeteer in each. Thus, we had three conditions: (1) Tom was controlled by Silicon Coppélia (n=27), (2) Human 1 controlled Tom (n=22), (3) Human 2 controlled Tom (n=5). Taken together, 27 participants interacted with an agent controlled by a human, and 27 participants interacted with an agent controlled by our software. This way, the behavior simulated by our model could be compared to behavior of the human puppeteers. In other words, this was an advanced kind of Turing Test where we compared the cognitive-affective structure between conditions. In a traditional Turing Test, participants do not know whether they interact with a computer or not whereas in our set-up participants were told they were interacting with a robot to avoid rejection of the dating partner on the basis of limited interaction possibilities.

2.4 Procedure

Participants were asked to take place behind a computer. They were instructed to do a speed-date session with an avatar. In the WOz, the human controlling the avatar was behind a wall, and thus invisible for the participants. After finishing the speed-dating session of about 10 minutes, the participants were asked to complete a questionnaire on the computer. After the experiment, participants in the WOz were debriefed that they were dating an avatar controlled by a human.

2.5 Measures

The questionnaire consisted of 97 Likert-type items with 0-5 rating scales, measuring agreement to statements. Together there were 15 scales. We designed five emotion scales for Joy, Anger, Hope, Fear, and Sadness, based on (Wallbot & Scherer, 1989). We also designed a scale for Situation Selection, with items such as 'Tom kept on talking about the same thing' and 'Tom changed the subject', and a scale for Affective Decision-Making, with items such as 'Tom followed his intuition' and 'Tom made rational choices'. For all eight parameters that were present in the I-PEFiC model (Ethics, Affordances, Similarity, Relevance, Valence, Involvement, Distance, Use Intentions), the questions from previous questionnaires (e.g., Van Vugt, Hoorn & Konijn, 2009) were adjusted and reused. However, because of the different application domain (i.e. speed dating), and because the questions were now about assessing how Tom perceived the participant, and not about how the participant perceived Tom, we found it important to check the consistency of these scales again.

A scale analysis was performed, in which items were removed until an optimal Cronbach's alpha was found and a minimum scale length of three items was achieved. If removing an item only increased Cronbach's alpha very little, the item was maintained. After scale analysis, a factor analysis was performed, to check divergent validity. After additional items were removed, again a scale analysis was performed (Appendix). All alphas, except those for Ethics and Similarity, were between .74 and .95. The scale for Similarity had an alpha of .66. Previous studies showed that the present Ethics scale was consistently reliable, and an important theoretical factor. Therefore, we decided to maintain the Ethics scale despite its feeble measurement quality.

2.6 Statistical Analyses

We performed a multivariate analysis of variance (MANOVA) on the grand mean scores to scales, to test whether the participants perceived a difference in Agent-type (software vs. human controlled). We performed paired t-tests for related groups of variables.

3 RESULTS

3.1 Emotions

To analyze the differences in perceived emotions in the three agent types, we performed a 3x5 multivariate analysis of variance (MANOVA) of the between-factor Agent-type (3: Silicon Coppélia, Human1, Human2) and the within-factor of Emotion (5: Joy, Sadness, Hope, Fear, Anger) on the grand mean scores to statements. The main effect of Agent-type on the grand mean scores to emotion scales was not significant ($F_{(2, 51)} = 1.68$, p < .196), whereas the main effect of the Emotion factor was significant (Pillai's Trace = .64, $F_{(4, 48)} = 21.59$, p < .001, $\eta^2_p = .64$). The interaction between Agent-type and Emotions was not significant (Pillai's Trace = .22, $F_{(8, 98)} = 1.545$, p < .152). More detailed results can be found in the Appendix.

Because the main effect of Agent-type to Emotion scales was not significant, this might mean that there was no effect of emotion at all within a condition. To check whether emotional behavior was diagnosed at all by the participants, we performed a one-sample t-test with 0 as the test value, equalling no emotions diagnosed. Results showed that all emotion scales differed significantly from 0. The smallest t-value was found for Anger ($t_{(2, 51)} = 8.777$, p < .001).

In addition, the significant main effect of the Emotion factor suggested that there were systematic differences in diagnosing emotions in Tom, which we analyzed by paired samples t-tests for all pairs of emotions. Out of the 10 thereby originated pairs, 6 pairs differed significantly. The 4 pairs that did not differ significantly were Joy and Hope (p < .444), Fear and Sadness (p < .054), Fear and Anger (p < .908), and Sad and Anger (p < .06). Joy (M = 3.05, SD = 1.03) and Hope (M = 2.96, SD = .82) were both recognized relatively much in Tom, whereas Fear (M=1.04, SD=.80), Sad (M=.84, SD=.66) and Anger (M=1.02, SD=.86) were recognized little in Tom.

In other words, the t-tests showed that emotions were recognized in all conditions, and the MANOVA showed that participants saw equal emotions in humans and robots alike.

3.2 Perceptions

To analyze the differences in perceived perceptions in the three agent-types, we performed a 3x8 MANOVA of the between-factor Agent-type (3: Silicon Coppélia, Human1, Human2) and the within-factor of Perception (8: Ethics, Affordances, Relevance, Valence, Similarity, Involvement, Distance, Use Intentions) on the grand mean scores to statements. The main effect of Agent-type on the perception scale scores was not significant (F < 1), whereas the main effect of the Perception factor was significant (Pillai's Trace = .87, $F_{(7, 43)} = 39.63$, p < .001, $\eta^2_p = .87$). The interaction between Agent-type and Perception was not significant (Pillai's Trace = .18, $F_{(14, 88)} = .635$, p < .828). More detailed results can be found in the Appendix.

Because the main effect of Agent-type to Perception scales was not significant, this might mean that there was no effect of perception at all within a condition. To check whether the perceptions of Tom were diagnosed at all by the participants, we performed a one-sample t-test with 0 as the test value, equalling no perceptions diagnosed. Results showed that all perception scales differed significantly from 0. The smallest t-value was found for Distance ($t_{(2, 51)} = 15.865$, p < .001).

In addition, the significant main effect of the Perception factor suggested that there were systematic differences in diagnosing perceptions in Tom, which we analyzed by paired samples t-tests for all pairs of perceptions. Out of the 28 thereby originated pairs, 23 pairs differed significantly. The pair that differed the most was Ethics and Distance (t(51) = 13.59, p < .001).

Tom's perceptions of Ethics (M = 3.86, SD = .68) and Affordances (M = 3.78, SD = .81) in the participant were rated the highest. His perceptions of feeling distant towards the participant (M = 1.77, SD = .93) were rated the lowest.

In other words, the t-tests showed that perceptions were recognized in all conditions, and the MANOVA showed that participants saw equal perceptions in humans and robots alike.

3.3 Decision-making Behavior

To analyze the differences in perceived decision-making behavior in the three agent-types, we performed a 3x2 MANOVA of the between-factor Agent-type (3: Silicon Coppélia, Human1, Human2) and the within-factor of Decision-making behavior (2: Affective decision making, Situation selection) on the grand mean scores to statements. The main effect of Agent-type was not significant (F < 1), whereas the main effect of Decision-making behavior was small but significant (Pillai's Trace = .088, $F_{(1, 51)} = 4.892$, p < .031, $\eta^2_p = .088$). The interaction between Agent-type and Decision-making behavior was not significant (Pillai's Trace

$= .04$, $F_{(2, 51)} = .1.128$, $p < .332$). More detailed results can be found in the Appendix.

Because the main effect of Agent-type to Decision-making behavior scales was not significant, this might mean that there was no effect of Decision-making behavior at all within a condition. To check whether decision-making behavior was diagnosed at all by the participants, we performed a one-sample t-test with 0 as the test value, equalling no decision-making behavior diagnosed. Results showed that both Situation selection ($t_{(2, 51)} = 14.562$, $p < .001$) and Affective decision-making ($t_{(2, 51)} = 15.518$, $p < .001$) both differed significantly from 0.

In addition, the significant main effect of the Perception factor on Agreement suggested that there were systematic differences in diagnosing perceptions in Tom, which we analyzed by paired samples t-test for affective decision-making (M = 2.24, SD = 1.07) and situation selection (M = 1.91, SD = 1.32). The pair differed significantly ($t(53) = 1.776$, $p < .081$).

In other words, the t-tests showed that decision-making behavior was recognized in all conditions, and the MANOVA showed that participants saw equal decision-making behavior in humans and robots alike.

4 DISCUSSION

4.1 Conclusions

In this paper, we equipped a virtual agent with Silicon Coppélia (Hoorn et al., 2012), a cognitive model of perception, affection, and affective decision-making. As an advanced, implicit version of a Turing Test, we let participants perform a speed-dating session with Tom, and asked them how they thought Tom perceived them during the speed-date. What the participants did not know, was that in one condition, a human was controlling Tom, whereas in the other condition, Tom was equipped with Silicon Coppélia.

A novel element in this experiment was that participants were asked to imagine how an agent perceived them. To our knowledge there does not exist previous research in which participants were asked to assess the perceptions of an artificial other. It is a nice finding, that the scales of I-PEFiC (Van Vugt et al., 2009), which were originally used to ask how participants perceived an interactive agent, could be used quite well to ask participants how they thought Tom perceived them.

The results showed that in this enriched and elaborated version of the classic Turing Test,

participants did not detect differences between the two versions of Tom. Not that the variables measured by the questionnaire did not have any effect; the effects just did not differ. Thus, within the boundaries of limited interaction possibilities, the participants felt that human and software perceived their moral fiber in the same way, deemed their relevance the same, and so on. The participants felt that human and software were equally eager to meet them again, and exhibited equal ways to select a situation and to make affective decisions. Also, the emotions the participants perceived in Tom during the speed-date session did not differ between conditions. Emotion effects could be observed by the participants, and these effects were similar for a human controlled avatar and software agent alike. This is good for the engineer who wants to use these models for application development, such as the design of virtual agents or robots. After all, on all kinds of facets, participants may not experience any difference between the expression of human behavior and behavior generated by our model.

4.2 Applications

Our findings can be of great use in many applications, such as (serious) digital games, virtual stories, tutor and advice systems, or coach and therapist systems. For example, Silicon Coppélia could be used to improve the emotional intelligence of a 'virtual crook' that could be used for police studies to practice situations in which the police officers should work on the emotions of the crook, for example questioning techniques (Hochschild, 1983). Another possible use of models of human processes is in software and/or hardware that interacts with a human and tries to understand this human's states and processes and responds in an intelligent manner. Many ambient intelligence systems (e.g., Aarts et al., 2001) include devices that monitor elderly persons. In settings where humans interact intensively with these systems, such as cuddle bots for dementia patients (e.g., Nakajima et al, 2001), the system can combine the data gathered from these devices with Silicon Coppélia to maintain a model of the emotional state of the user. This can enable the system to adapt the type of interaction to the user's needs.

Silicon Coppélia can also be used to improve self-help therapy. Adding the moral reasoning system will be very important for that matter. Humans with psychological disorders can be supported through applications available on the Internet and virtual communities of persons with

similar problems.

New communication technologies have led to an impressive increase of self-help programs that are delivered through the Internet (e.g., Spek et al., 2007). Several studies concluded that self-help therapies can be more efficient in reducing mental health problems, and less expensive than traditional therapy (e.g., Andrews et al., 2001; Bijl and Ravelli, 2000; Cuijpers, 1997; Spek et al., 2007).

Web-based self-help therapy can be a solution for people who would otherwise not seek help, wishing to avoid the stigma of psychiatric referral or to protect their privacy (Williams, 2001). The majority of persons with a mental disorder in the general population do not receive any professional mental health services (an estimated 65%) (Andrews et al., 2001; Bijl and Ravelli, 2000). In many occupations, such as the police force, the fire service and farming, there is much stigma attached to receiving psychological treatment, and the anonymity of Web-based self-help therapy would help to overcome this. Also many other people feel a barrier to seek help for their problems through regular health-care systems; e.g., in a study by Spek et al. (2007) about internet-based cognitive behavioral therapy for sub-threshold depression for people over 50 years old, many participants reported not seeking help through regular health-care systems because they were very concerned about being stigmatized. Patients may be attracted to the idea of working on their own to deal with their problems, thereby avoiding the potential embarrassment of formal psychotherapy (Williams, 2001).

Further, self-help therapy is particularly suited to remote and rural areas, where ready access to a face-to-face therapist cannot be economically justified. Self-help therapy may also be useful in unusual environments such as oilrigs and prisons, where face-to-face therapy is not normally available. Self-help therapy can also be offered to patients while they are on a waiting list, with the option to receive face-to-face therapy later, if required (Peck, 2007)

Self-help therapy may be even more successful when the interface is enhanced or replaced by a robot therapist that has Silicon Coppélia installed. The anonymity of robot-supported self-help therapy could overcome potential embarrassment of undergoing formal treatment. When regular therapy puts up too high a threshold, a robot therapist is less threatening, what the patient reveals is inconsequential, the patient is in control, and all in all, interaction with the virtual therapist has a "dear diary" effect. As if you were speed-dating with a real partner.

4.3 Future Research

In future research, we will test an extended version of the current model, using robots in the healthcare domain. So-called Caredroids will play a chess game with the patient as a form of daytime activity. Based on whether the agent reaches its goals (winning and losing when the agent has ambitions to win or lose), the likelihood of these goals, and the expectedness of the move of the user and the outcome of a game, the emotions joy, distress, hope, fear and surprise are simulated and shown by the agent by means of bodily expressions. The Caredroid will be able to trade rational choices to win the game for affective choices to let the human opponent win if she is nice to him.

Additionally, we will integrate Silicon Coppélia with a moral reasoning system that can solve medical ethical dilemmas (Pontier and Hoorn, 2012). In this system, actions are evaluated against a number of moral principles to point out ethical dilemmas in employing robot care.

In entertainment settings, we often like characters that are naughty; the good guys often are quite boring (Konijn and Hoorn, 2005). In Silicon Coppélia (Hoorn et al., 2012), this could be implemented by updating the affective decision making module. Morality would be added to the other influences that determine the Expected Satisfaction of an action in the decision making process. By doing so, human affective decision-making behavior could be further explored. Some inital steps in doing this were taken in (Pontier, Widdershoven and Hoorn, 2012).

ACKNOWLEDGEMENTS

REFERENCES

Aarts, E., Harwig, R., and Schuurmans, M., 2001. Ambient Intelligence. In *The Invisible Future: The Seamless Integration of Technology into Everyday Life*, McGraw-Hill, 2001.

Andrews, G., Henderson, S., & Hall, W., 2001. Prevalence, comorbidity, disability and service utilisation. Overview of the Australian National

Mental Health Survey. In *The British Journal of Psychiatry*, Vol. 178, No. 2, pp. 145-153.

Barrett, L. F., Robin, L., Pietromonaco, P. R., Eyssell, K. M., 1998. Are Women The "More Emotional Sex?" Evidence from Emotional Experiences in Social Context. In *Cognition and Emotion*, 12, pp. 555-578.

Bijl, R. V., & Ravelli, A., 2000. Psychiatric morbidity, service use, and need for care in the general population: results of The Netherlands Mental Health Survey and Incidence Study. In *American Journal of Public Health*, Vol. 90, Iss. 4, pp. 602-607.

Bosse, T., Gratch, J., Hoorn, J. F., Pontier, M. A., Siddiqui, G. F., 2010. 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.

Bosse, T., Pontier, M. A., Treur, J., 2007. A Dynamical System Modelling Approach to Gross´ Model of Emotion Regulation. In *Proceedings of the 8th International Conference on Cognitive Modeling*, ICCM'07, pp. 187-192, Taylor and Francis.

Brave, S., Nass, C., & Hutchinson, K., 2005. Computers that care: investigating the effects of orientation of emotion exhibited by an embodied computer agent. *International Journal of Human-Computer Studies*, 62(2), pp. 161–178.

Cuijpers, P., 1997. Bibliotherapy in unipolar depression, a meta-analysis. In *Journal of Behavior Therapy & Experimental Psychiatry*, Vol. 28, No. 2, pp. 139-147

Gross, J. J., 2001. Emotion Regulation in Adulthood: Timing is Everything. In *Current Directions in Psychological Science*, 10(6), pp. 214-219.

Hochschild, A. R., 1983. *The managed heart*. Berkeley: University of California Press.

Hoorn, J. F., Pontier, M. A., Siddiqui, G. F., 2008. When the user is instrumental to robot goals. First try: Agent uses agent. In *Proceedings of International Conference on Web Intelligence and Intelligent Agent Technology*, WI-IAT '08, IEEE/WIC/ACM, Sydney AU, 296-301.

Hoorn, J. F., Pontier, M. A., & Siddiqui, G. F., 2012. Coppélius' Concoction: Similarity and Complementarity Among Three Affect-related Agent Models. *Cognitive Systems Research*, 2012, pp. 33-49.

Konijn, E. A., & Hoorn, J. F., 2005. Some like it bad. Testing a model for perceiving and experiencing fictional characters. *Media Psychology*, 7(2), 107-144.

Konijn, E. A., & Van Vugt, H. C., 2008. Emotions in mediated interpersonal communication. Toward modeling emotion in virtual agents. In Konijn, E.A., Utz, S., Tanis, M., & Barnes, S. B. (Eds.). *Mediated Interpersonal Communication*. New York: Taylor & Francis group/ Routledge.

Landauer, T. K., 1987. Psychology as a mother of invention. *Proceedings of the SIGCHI/GI conference on Human factors in computing systems and graphics interface*, pp. 333-335.

Marsella, M., Gratch, J., 2009. EMA: A Process Model of Appraisal Dynamics. In *Journal of Cognitive Systems Research*, Vol. 10, No. 1, pp. 70-90.

Nakajima, K., Nakamura, K., Yonemitsu, S., Oikawa, D., Ito, A., Higashi, Y., Fujimoto, T., Nambu, A., Tamura, T., 2001. Animal-shaped toys as therapeutic tools for patients with severe dementia. Engineering in Medicine and Biology Society. In *Proceedings of the 23rd IEEE Conference*, vol. 4, 3796-3798.

Peck, D., 2007. Computer-guided cognitive–behavioral therapy for anxiety states. *Emerging areas in Anxiety,* Vol. 6, No. 4, pp. 166-169.

Picard, R. W., 1997. *Affective Computing*. Cambridge, MA: MIT Press.

Pontier, M. A., & Hoorn, J. F., 2012. Toward machines that behave ethically better than humans do. In *Proceedings of CogSci'12*, pp. 2198-2203.

Pontier, M. A., Siddiqui, G. F., 2009. Silicon Coppélia: Integrating Three Affect-Related Models for Establishing Richer Agent Interaction. In *WI-IAT'09, International Conference on Intelligent Agent Technology*, vol. 2, pp. 279—284.

Pontier, M. A., Siddiqui, G. F., & Hoorn, J., 2010. Speed-dating with an Affective Virtual Agent - Developing a Testbed for Emotion Models In: J. Allbeck et al. (Eds.), In IVA'10, *Lecture Notes in Computer Science*, Vol. 6356, pp. 91-103.

Pontier, M. A., Widdershoven, G. A. M., and Hoorn, J. F., 2012. Moral Coppélia - Combining Ratio with Affect in Ethical Reasoning. In: *Proceedings of the 13th Ibero-American Conference on Artificial Intelligence*, IBERAMIA'12, in press.

Smith, C. A., & Lazarus, R. S., 1990. Emotion and adaptation. In L. A. Pervin (Ed.), *Handbook of personality: Theory & research*, (pp. 609-637). NY: Guilford Press.

Spek, V., Cuijpers, P., Nyklíĉek, I., Riper, H., Keyzer, J., & Pop, V., 2007. Internet-based cognitive behavior therapy for emotion and anxiety disorders: a meta-analysis. In *Psychological Medicine*, Vol. 37, pp. 1-10

Turing, A. M., 1950. Computing Machinery and Intelligence. *Mind* 54(236), pp. 433-460.

Van Vugt, H. C., Hoorn, J. F., & Konijn, E. A., 2009. Interactive engagement with embodied agents: An empirically validated framework. In *Computer Animation and Virtual Worlds*, Vol. 20*,* pp. 195-204

Wallbot, H. G., & Scherer, K. R., 1989. Assessing Emotion by Questionnaire. *Emotion Theory, Research and Experience,* Vol. 4, pp. 55-82.

Williams, C., 2001. Use of Written Cognitive-Behavioral Therapy Self-Help Materials to treat depression. In *Advances in Psychiatric Treatment*, Vol. 7, pp. 233–240.

APPENDIX.

http://camera-vu.nl/matthijs/SpeeddateAppendix.pdf