

A Robot's Experience of Another Robot: Simulation

Tibor Bosse¹, Johan F. Hoorn², Matthijs Pontier^{1,2}, and Ghazanfar F. Siddiqui^{1,2}

¹VU University, Department of Artificial Intelligence
De Boelelaan 1081, 1081 HV Amsterdam, The Netherlands
{tbosse, mpr210, ghazanfa}@few.vu.nl

²VU University, Center for Advanced Media Research Amsterdam
Buitenveldertselaan 3, 1082 VA Amsterdam, The Netherlands
JF.Hoorn@fsw.vu.nl

Abstract

To develop a robot that is able to recognize and show affective behavior, it should be able to regulate simultaneously occurring tendencies of positive and negative emotions. To achieve this, the current paper introduces a computational model for involvement-distance trade-offs, based on an existing theoretical model. The main mechanisms of this model have been represented as regression equations, using the LEADSTO modeling environment. A number of simulation experiments have been performed, which indicated that the model is adequate for simulating the dynamics of involvement-distance trade-offs and their influence on satisfaction. More specifically, the experiments confirmed the empirical finding that positive features do not exclusively increase involvement.

Keywords: emotions, computational modeling, involvement-distance trade-off, simulation.

Introduction

We are a research group on a mission. Our aim is to model aspects of emotion regulation and involvement-distance trade-offs to build a robot that can contribute to the wellbeing of patients in need of psychological support. Once the software is capable of simulating emotion-regulating mechanisms and, where appropriate, can trade involvement for distance (or vice versa), we will develop a prototype virtual therapist that is tested against real patients. This therapist should be capable of recognizing emotional behavior and should respond to that in an emotionally appropriate way. People often find it hard to admit that they are in need of therapy or coaching. Experimenting with a virtual therapist as a support tool for self-diagnosis may help to overcome that barrier more easily.

However, as luring as this far horizon may be, there is quite some groundwork to be done first and this paper is addressing some of the modeling issues involved. In a counterpart paper in this volume (Hoorn, 2008), we dealt with theoretical matters of humans perceiving a virtual other and the way to formalize this (i.e. fuzzy sets). The idea was that the empirically validated models of human encounters with agents (e.g., Van Vugt et al., 2007) and models of emotion regulation (e.g., Gross, 1998; 2001; Marsella and Gratch, 2003) could be integrated and used to do the reverse, have a robot determine its level of engagement with

its user and have it choose the appropriate affective response to it.

There are two things we want to do in this paper, leaving aside many of the other important issues (Hoorn, 2008). First, we want to test the formula of Werners (1988) to see whether it represents the trade-off well between involvement (the robot becomes friendly with its user) and distance (the robot stays aloof). This involvement-distance trade-off is the fundamental mechanism in user encounters with agents (e.g., Van Vugt et al., 2007), film characters (Konijn and Hoorn, 2005), and photographs of people (Konijn and Bushman, 2007). Although one would expect the notions of “involved” (i.e., the tendency to be friendly) and “at a distance” (i.e., the tendency to stay aloof) to be part of one and the same continuum, the above studies have pointed out that their interplay is in reality more subtle. For example, humans can at the same time find another person attractive but morally repulsive, useful to achieve certain goals but distasteful in his or her manners (e.g., Konijn and Hoorn, 2005). We want to see whether Werners’ fuzzy trade-off works well enough to implement this mechanism in our prototype “emobot” (cf. Hooley et al., 2004).

Second, we want to show that we can simulate the way perceptual and experiential factors feed into the involvement-distance trade-off. A number of factors help establish the involvement-distance trade-off. For the body of empirical work that sustains this view, see Hoorn (2008). The *ethics* of the user, for example, whether the user is of good or bad intent (take care of or kill the robot) directly affects how the trade-off develops. Another important factor is the *affordances* a user offers as an aid or obstacle for task performance. A user that performs a simple task causes less trouble for the machinery of the robot than a demanding user does. Additional factors are the *aesthetics* of the user (beautiful or ugly), *epistemics* or realism (realistic or unrealistic), and *similarity* (user resembles robot or not). The upshot is that not anything positive leads to involvement but can increase distance as well (Van Vugt, Hoorn et al., 2006; Van Vugt, Konijn et al., 2006). This has to do with the goals of the robot. If it admires the skills of a user but if those same skills mean it is going to be dumped soon, being skilled raises involvement - but for different reasons - distance as well. This redistribution of information runs via the factors of *relevance* (whether user is important

to achieve robot goals such as being needed or timely maintenance), and *valence* (the expected positive or negative results of interacting with the user). For an overview of the complete model, see (Van Vugt, Hoorn et al., 2006), Figure 1.

The involvement-distance trade-off that is fed by all these factors is used for *affective response selection* (see Hoorn, 2008). This is the process of (qualitatively) evaluating the emotional significance of events, see (Gratch, 2000; Marsella and Gratch, 2003). In this paper, we focus explicitly on events related to what Gross (2001) calls *situation selection*, i.e. selecting situations that bring the robot in a more desirable state. Examples are walking away from a person the robot feels uncomfortable with and looking for another conversation partner. Thus, in our model the term satisfaction indicates the level of appreciation that a robot attaches to a certain situation it is in. If this level of current satisfaction is too low, the robot may want to select another situation based on the expected satisfaction in the future situation.

To account for the trade-off between conflicting options (e.g., ‘involvement’ vs. ‘distance’) that influence the level of satisfaction, Werners (1988, p. 297) employs the *fuzzy_AND*-operator γ . Following this approach, each feature u in a trade-off has a membership function μ in the fuzzy sets of involvement (\tilde{I}) and distance (\tilde{D}), which allows the feature to move between the minimum and maximum degree of membership to these sets:

$$\mu_{\tilde{I}} \text{ and } \mu_{\tilde{D}} (\mu_{\tilde{I}}(u), \mu_{\tilde{D}}(u)) = \gamma \cdot \min\{\mu_{\tilde{I}}(u), \mu_{\tilde{D}}(u)\} + ((1 - \gamma)(\mu_{\tilde{I}}(u) + \mu_{\tilde{D}}(u)) / n),$$

where $u \in U$, $\gamma \in [0, 1]$, and n is the number of fuzzy sets (\tilde{I} and \tilde{D}) for which the mean is calculated¹. Basically, then, there are two ways to calculate a trade-off, using the ($\gamma \cdot \min$) option or the ($\gamma \cdot \max$):

$$\gamma \cdot \max\{\mu_{\tilde{I}}(u), \mu_{\tilde{D}}(u)\} + ((1 - \gamma)(\mu_{\tilde{I}}(u) + \mu_{\tilde{D}}(u)) / n)$$

When the robot feels ambivalent about its user (“sympathetic but clumsy”), using either option may lead to quite different results for the level of satisfaction. When the mean of involvement and distance (the part after $(1 - \gamma)$ in the formula) is the same, the ($\gamma \cdot \min$) version favors decision options in which the involvement and distance values are close to each other, i.e., to decision options which involve relatively more doubt. The ($\gamma \cdot \max$) version favors options in which involvement and distance differ more from each other, i.e., options of which the robot feels less ambivalent.

Our hypothesis (H1), then, explores whether our system can simulate counter-intuitive empirical results concerning the influence of features on involvement and distance:

H1: Positive features do not necessarily and exclusively increase involvement. Due to the redistribution of information via relevance and valence, also distance can be increased. (The same mechanism may apply to negative features that partly increase involvement).

Simulation Model

To simulate the dynamics of involvement-distance trade-offs and their influence on satisfaction, the theoretical model by Hoorn (2008) was taken as a basis and was implemented in the LEADSTO modeling environment (Bosse et al., 2007). This environment enables the modeler to represent the most elementary steps of a process in terms of direct temporal dependencies, and features a dedicated simulation tool. In this section, we describe how we represented the basic mechanisms of the model in LEADSTO². First, a number of design decisions had to be made. In particular, we chose to treat what are actually factor levels as single features. We then represented the features of different agents (e.g., their goodness, realism, or beauty) by real numbers between 0 and 1. In addition, the satisfaction an agent had in a particular situation was represented by a real number between 0 and 1. To model the impact of (the perception of) the different features on an agent’s satisfaction, Hoorn (2008) proposed to use fuzzy set theory. The idea of fuzzy set theory is that features have membership functions for various sets, which determine to what degree they are member of these sets. Within LEADSTO, this principle can easily be simulated by using regression equations, but only to the extent that we used factor levels for features. The process of extracting factor levels from features would require a more elaborated model, which is beyond the scope of this paper.

Domain

We created a virtual environment that was inhabited by a number of virtual agents. These agents are fans of soccer teams and express this by wearing club clothes of their favorite team. When these agents meet, to a certain degree they are involved with and at a distance towards each other. These tendencies are based on features of the agents, according to the formulas described in this section. Table 1 shows certain variables that were used in these formulas.

Table 1: Variable names and meanings.

| Variable | Meaning | Range |
|--|---|--------|
| Perceived _(<Feature>, A1, A2) | Agent1’s perception of a certain feature of Agent2 | [0, 1] |
| Designed _(<Feature>, A2) | Value assigned by ‘the designer’ to a certain feature of Agent2 | [0, 1] |
| Bias _(A1, A2, <Feature>) | Bias that Agent1 has about a certain feature of Agent2 | [0, 2] |
| Skill _(A1, <Language>) | Skill of Agent1 in a certain language | [0, 1] |
| ExpectedSkill _(A1, A2, language) | Skill Agent1 expects Agent2 to have in a certain language | [0, 1] |
| $\beta_{\text{factor1} \leftarrow \text{factor2}}$ | Regression weight factor2 has for another factor1 for a certain agent | [0, 1] |
| $\gamma_{\text{inv-dist}}$ | Variable that is used to calculate the involvement-distance trade-off | [0, 1] |
| Satisfaction _(A1, A2) | Indicates to what extent Agent1 is satisfied with Agent2 | [0, 1] |

¹ In the remainder of this paper, we only address the simple case of $n=2$.

² For details about the model, see part A of the appendix.

Aesthetics

Each agent has a value for ‘*designed* beautiful / ugly’. This is a value the designer expects to raise in the user, or in another agent, based on general principles of aesthetics. This value could be seen as the mean ‘score’ an agent receives for its beauty / ugliness from all other agents. This *designed* value has a data-driven influence on how agents perceive the beauty of another agent. The variable *bias* represents the concept-driven influence on how agents perceive other agents’ beauty. Depending on its value, a bias may increase or decrease an agent’s perception of another agent’s beauty. Note that ‘another agent’ could also be the agent itself. This is represented by the following formulas, given in mathematical format.

$$\text{Perceived}_{(\text{Beautiful}, A1, A2)} = \text{Bias}_{(A1, A2, \text{Beautiful})} * \text{Designed}_{(\text{Beautiful}, A2)}$$
$$\text{Perceived}_{(\text{Ugly}, A1, A2)} = \text{Bias}_{(A1, A2, \text{Ugly})} * \text{Designed}_{(\text{Ugly}, A2)}$$

When two agents meet for the first time, they will assign a *perceived value* in the range [0, 1] to each other’s beauty and ugliness according to the formulas above. *Bias* in the range [0, 2] is multiplied with the designed value for the feature in the range [0, 1]. If agent A1 has a *bias* of 1 for, for instance, the beauty of agent A2, then A1 does not under- or overestimate the beauty of A2. If the *bias* is bigger than 1, then A1 is relatively positive about the beauty of agent A2. When the resulting value for the perceived feature is bigger than 1, it is set to 1, to prevent the formula from going out of range.

Ethics

In line with soccer tradition, good guys are those who are fan of the club, and bad guys are fan of the opponent. Agents recognize good and bad guys by the club clothes they are wearing. E. g., if agent A1 is a fan of the soccer club Ajax, and agent A2 wears club clothes of Ajax, then A1 will think of A2 as a ‘good guy’, but if A2 wears club clothes of the rival soccer club Feyenoord, then A1 will think of A2 as a ‘bad guy’. This is represented by:

$$\text{Perceived}_{(\text{Good}, A1, A2)} = \text{Satisfaction}_{(A2, \text{Club})}$$
$$\text{Perceived}_{(\text{Bad}, A1, A2)} = 1 - \text{Satisfaction}_{(A2, \text{Club})}$$

These formulas say that when two agents meet for the first time, they will perceive a value in the range [0, 1] for each other’s goodness and badness. The perceived goodness is exactly the same value as the level of satisfaction the agent attaches to the club of which the other agent is a fan. The perceived badness is 1 minus the level of satisfaction. If an agent wears neutral clothes, the values of good and bad are assigned according to a variable that reflects the agents’ perception of neutral clothes (which is 0.3 for both good and bad in the simulations in this paper).

Epistemics

The first time agents meet, each agent perceives the epistemics (or realism) of itself and other agents, the same way it perceives the aesthetics of itself and other agents:

$$\text{Perceived}_{(\text{Realistic}, A1, A2)} = \text{Bias}_{(A1, A2, \text{Realistic})} * \text{Designed}_{(\text{Realistic}, A2)}$$
$$\text{Perceived}_{(\text{Unrealistic}, A1, A2)} = \text{Bias}_{(A1, A2, \text{Unrealistic})} * \text{Designed}_{(\text{Unrealistic}, A2)}$$

Affordances

In the simulation model, the languages Urdu, English, and Dutch are used as the affordances to have a conversation about soccer. Each agent has a certain *skill* level for each language. Agents perceive each other’s affordances according to the expectations they have about the possibilities to communicate with the other agent, according to the following formulas (where the sum over all languages is taken):

$$\text{Perceived}_{(\text{Aid}, A1, A2)} = \sum(\text{ExpectedSkill}_{(A1, A2, \text{language})} * \text{Skill}_{(A1, \text{language})})$$
$$\text{Perceived}_{(\text{Obstacle}, A1, A2)} = 1 - \sum(\text{ExpectedSkill}_{(A1, A2, \text{language})} * \text{Skill}_{(A1, \text{language})})$$

When two agents meet, they assign a value to each other’s affordances (aid and obstacle) in the range [0, 1], using the presuppositions they have about the language skills of the other agent, which is based on outer appearance. E.g., when Agent A2 has a dark skin, in the simulation, Agent A1 will think Agent A2 has good skills in Urdu, average skills in English, and bad skills in Dutch. Because of this, the value of *aid* is calculated by taking the sum of the language skills of Agent A1 multiplied by the language skills A1 expects A2 to have. These expectations of Agent A1 about the language skills of Agent A2 are normalized, and are based on skin color (although politically incorrect, this was convenient for simulation purposes). The perceived value for *obstacle* was 1 minus the calculated value for *aid*.

Similarity

For an agent to perceive its similarity with another agent, it needs to perceive the features of the self. Agents perceive their own features the same way they perceive the aesthetics and epistemics of other agents. Only this time, the *bias* is the bias in self-perception, instead of in the perception of another agent.

$$\text{Perceived}_{(\text{Feature}, A1, A1)} = \text{Bias}_{(A1, A1, \text{Feature})} * \text{Designed}_{(\text{Feature}, A1)}$$

Similarity is perceived according to the differences between the agent’s perception of its own features (*good*, *bad*, *beautiful*, *ugly*, *realistic* and *unrealistic*) and its perception of the features of the other agent (where the sum over ranges over these six features):

$$\text{Similarity}_{(A1, A2)} = 1 - (\sum(\beta_{\text{sim}_{\text{feature}}} * \text{abs}(\text{Perceived}_{(\text{Feature}, A1, A2)} - \text{Perceived}_{(\text{Feature}, A1, A1)})))$$
$$\text{Dissimilarity}_{(A1, A2)} = \sum(\beta_{\text{ds}_{\text{feature}}} * \text{abs}(\text{Perceived}_{(\text{Feature}, A1, A2)} - \text{Perceived}_{(\text{Feature}, A1, A1)}))$$

To calculate the dissimilarity between two agents, the differences between the perceived values for its own features, and those perceived for the other agent are taken. These differences are all added, with a certain (regression) weight β . Similarity is calculated in a similar manner, but with different weights, and 1 was subtracted by the sum of all differences.

Relevance, Valence, Involvement and Distance

The formulas in this paragraph were designed using generalized linear models (Nelder & Wedderburn, 1972; McCullagh & Nelder, 1983). Hoorn (2008) shows that the calculated dependent variable (e.g., relevance) is fed by a number of contributing variables. Each contributing variable

has a certain main effect on the dependent variable. The size of this main effect is represented by a (regression) weight β , the same way as for calculating similarity. When two variables interact, the interaction effect on the dependent variable is calculated by multiplying the product of the values of these two variables with a certain regression weight, which accounts for the interaction effect on the dependent variable. When the interaction is over-additive, the weight will be positive, and when it is under-additive, the weight will be negative.

The formula for the calculation of a variable A that is dependent on the variables B, C, and D, of which C and D interact, would be: $A = \beta_B * B + \beta_C * C + \beta_D * D + \beta_{CD} * C * D$. In this formula, β_B , β_C , and β_D are the (regression) weights for the main effect of variables B, C, and D respectively, and β_{CD} is the (regression) weight for the interaction effect of C and D.

Due to space limitations, the formulas for relevance, valence, involvement and distance are not given completely, but all the effects on the variables are summarized in Table 2. The formulas are then constructed using the algorithm described above. For theoretical reasons, each variable in Table 2 is in the actual formula split up in two unipolar variables (ethics is split up into *good* and *bad*, valence is split up into *positive valence* and *negative valence*, engagement is split up into *involvement* and *distance*, etc.).

Table 2: Effects of features on relevance, valence, involvement and distance.

| Effects on: | Main effects | Interaction effects |
|-------------|---|--|
| Relevance | Ethics Aesthetics Epistemics Affordances | Ethics x Affordances Ethics x Aesthetics x Epistemics |
| Valence | Ethics Aesthetics Epistemics Affordances | Ethics x Affordances Ethics x Aesthetics x Epistemics |
| Engagement | Similarity Relevance Valence | Relevance x Valence |

For example, the formula to calculate relevance looks as follows (where ‘Perceived’ has been abbreviated as ‘Perc’).

$$\begin{aligned}
\text{Relevance}_{(A1, A2)} = & \beta_{\text{rel}_{\leftarrow \text{good}}} * \text{Perc}_{(\text{Good}, A1, A2)} + \\
& \beta_{\text{rel}_{\leftarrow \text{bad}}} * \text{Perc}_{(\text{Bad}, A1, A2)} + \\
& \beta_{\text{rel}_{\leftarrow \text{bea}}} * \text{Perc}_{(\text{Beautiful}, A1, A2)} + \\
& \beta_{\text{rel}_{\leftarrow \text{ugly}}} * \text{Perc}_{(\text{Ugly}, A1, A2)} + \\
& \beta_{\text{rel}_{\leftarrow \text{real}}} * \text{Perc}_{(\text{Realistic}, A1, A2)} + \\
& \beta_{\text{rel}_{\leftarrow \text{unr}}} * \text{Perc}_{(\text{Unrealistic}, A1, A2)} + \\
& \beta_{\text{rel}_{\leftarrow \text{aid}}} * \text{Perc}_{(\text{Aid}, A1, A2)} + \\
& \beta_{\text{rel}_{\leftarrow \text{obst}}} * \text{Perc}_{(\text{Obstacle}, A1, A2)} + \\
& \beta_{\text{rel}_{\leftarrow \text{good-aid}}} * \text{Perc}_{(\text{Good}, A1, A2)} * \text{Perc}_{(\text{Aid}, A1, A2)} + \\
& \beta_{\text{rel}_{\leftarrow \text{good-obst}}} * \text{Perc}_{(\text{Good}, A1, A2)} * \text{Perc}_{(\text{Obstacle}, A1, A2)} + \\
& \beta_{\text{rel}_{\leftarrow \text{bad-obst}}} * \text{Perc}_{(\text{Bad}, A1, A2)} * \text{Perc}_{(\text{Obstacle}, A1, A2)} + \\
& \beta_{\text{rel}_{\leftarrow \text{bad-aid}}} * \text{Perc}_{(\text{Bad}, A1, A2)} * \text{Perc}_{(\text{Aid}, A1, A2)} + \\
& \beta_{\text{rel}_{\leftarrow \text{good-bea-real}}} * \text{Perc}_{(\text{Good}, A1, A2)} * \text{Perc}_{(\text{Beautiful}, A1, A2)} * \text{Perc}_{(\text{Realistic}, A1, A2)} + \\
& \beta_{\text{rel}_{\leftarrow \text{good-bea-unr}}} * \text{Perc}_{(\text{Good}, A1, A2)} * \text{Perc}_{(\text{Beautiful}, A1, A2)} * \text{Perc}_{(\text{Unrealistic}, A1, A2)} + \\
& \beta_{\text{rel}_{\leftarrow \text{good-ugly-real}}} * \text{Perc}_{(\text{Good}, A1, A2)} * \text{Perc}_{(\text{Ugly}, A1, A2)} * \text{Perc}_{(\text{Realistic}, A1, A2)} + \\
& \beta_{\text{rel}_{\leftarrow \text{good-ugly-unr}}} * \text{Perc}_{(\text{Good}, A1, A2)} * \text{Perc}_{(\text{Ugly}, A1, A2)} * \text{Perc}_{(\text{Unrealistic}, A1, A2)} + \\
& \beta_{\text{rel}_{\leftarrow \text{bad-bea-real}}} * \text{Perc}_{(\text{Bad}, A1, A2)} * \text{Perc}_{(\text{Beautiful}, A1, A2)} * \text{Perc}_{(\text{Realistic}, A1, A2)} + \\
& \beta_{\text{rel}_{\leftarrow \text{bad-bea-unr}}} * \text{Perc}_{(\text{Bad}, A1, A2)} * \text{Perc}_{(\text{Beautiful}, A1, A2)} * \text{Perc}_{(\text{Unrealistic}, A1, A2)} + \\
& \beta_{\text{rel}_{\leftarrow \text{bad-ugly-real}}} * \text{Perc}_{(\text{Bad}, A1, A2)} * \text{Perc}_{(\text{Ugly}, A1, A2)} * \text{Perc}_{(\text{Realistic}, A1, A2)} + \\
& \beta_{\text{rel}_{\leftarrow \text{bad-ugly-unr}}} * \text{Perc}_{(\text{Bad}, A1, A2)} * \text{Perc}_{(\text{Ugly}, A1, A2)} * \text{Perc}_{(\text{Unrealistic}, A1, A2)}
\end{aligned}$$

Satisfaction

Within our model, satisfaction is a certain appreciation the agents attach to the possible decisions they can make (about situation selection). They use their expected satisfaction with each option, to decide which option ‘feels’ best for them. Satisfaction is calculated by a trade-off between involvement and distance:

$$\begin{aligned}
\text{Satisfaction}_{(A1, A2)} = & \gamma_{\text{inv-dist}} * \max(\text{Involvement}_{(A1, A2)}, \text{Distance}_{(A1, A2)}) + \\
& (1 - \gamma_{\text{inv-dist}}) * ((\text{Involvement}_{(A1, A2)}, \text{Distance}_{(A1, A2)}) / n)
\end{aligned}$$

When there is relatively more involvement, this will lead to a relatively more positive type of approach towards the other agent. Note that a lot of distance also can lead to a high satisfaction, reflecting a desire for a negative approach (“Beat up the soccer opponent”).

The trade-off is calculated using a variant of the fuzzy_AND-operator γ (Werners, 1988; Zimmermann, 1994). In the simulation experiments, two variants of this formula tested H1. In the min version, a part γ was taken of the minimum of *involvement* and *distance*, and a part $(1 - \gamma)$ was taken of the mean of *involvement* and *distance*, as originally proposed by Werners. In the max version, instead of a part γ of the minimum, a part γ of the maximum of *involvement* and *distance* was taken. In this paper, the value for γ is always set to 0.5.

Simulation Results

To test our hypothesis, the simulation model introduced in the previous section was used to perform a number of experiments under different parameter settings. In each experiment, three agents were involved, named Harry, Barry, and Gary. The results of these experiments are described below. Due to space limitations, not all parameter settings are shown in this paper³.

Baseline Condition. To start, an initial experiment was performed, to test whether the model behaves as expected, and as a control condition for the remaining experiments. In this condition, all the agents had a white skin, and wore Ajax clothes. For the chosen regression weights, see part B of the appendix (to give one example, $\beta_{\text{pv}_{\leftarrow \text{good}}} = 0.3$, which means the regression weight of *good* on *positive valence* is 0.3). Because all language skills for each agent were set to 0, which means they had no language skills at all, they expected not to be able to communicate with each other, which resulted in assigning 1 to *obstacle*, and 0 to *aid* for each other. With the formula for calculating *good* and *bad*, ‘good = 1’ implies ‘bad = 0.’ For this reason, the appraisal the agents attached to Ajax were all set to 0.5, which resulted in all agents assigning 0.5 to each other’s goodness, as well as their badness. The satisfaction of each agent with Feyenoord and all $\text{Designed}_{\langle \text{Feature} \rangle}$ parameters for the agents were set to 0. All *bias* parameters were set to the neutral value of 1 for each agent. For all agents these parameter settings led to an *involvement* of 0.11, a *distance* of 0.25,

³ For a detailed description of parameter settings, see appendix, part B.

and a level of *satisfaction* with each other of 0.21. These values are identical, because all agents are identical. This situation functions as a baseline for the following experiments.

Next, in order to test whether our system could simulate counter-intuitive empirical results (H1) concerning the influence of features on involvement and distance, we experimented with changing the values of *aesthetics* and *epistemics*, see Experiment 1 and 2.

Experiment 1: Aesthetics – beautiful vs. ugly. In this experiment, the parameter settings were the same as in the baseline condition, except that in this experiment, Barry was beautiful (beautiful = 1), and Gary was ugly (ugly = 1). Because of this, Harry's involvement towards Barry (0.11→0.18) increased. Surprisingly, also his distance towards Barry (0.25→0.31) increased. Moreover, both Harry's involvement (0.11→0.14) and distance (0.25→0.37) towards Gary increased as well. It is clear that increasing the value for *beautiful* adds relatively more to *involvement*, and increasing the value for *ugly* adds relatively more to *distance*. As *beautiful* is a positive feature, which would intuitively be expected to *only* increase involvement, and *ugly* is a negative feature, which would intuitively be expected to *only* increase distance, this corresponds with H1.

Experiment 2: Epistemics – realistic vs. unrealistic. In this experiment, the parameter settings were the same as in the baseline condition, except that in this experiment, Barry was realistic (realistic = 1), and Gary was unrealistic (unrealistic = 1). Because of this, Harry's involvement towards Barry (0.11→0.14) increased. Surprisingly, also his distance towards Barry (0.25→0.30) increased. Moreover, both Harry's involvement (0.11→0.14) and distance (0.25→0.31) towards Gary increased as well. As a result of the chosen regression weights, these effects were much smaller than the effects of adding *beautiful* and *ugly*. Adding to *realistic* adds relatively more to *involvement*, and adding *unrealistic* adds relatively more to *distance*, although this is much less clear than the difference between adding *beautiful* and *ugly*. Because *realistic* is a positive feature, which traditionally is expected to *only* increase involvement, and *unrealistic* is a negative feature, which in conventional theories would *only* increase distance, this result confirms H1.

Additional Experiments

In addition to the above experiments, we experimented with changing the values of *ethics* and *affordances*. However, within these formulas, 'good = 1' implies 'bad = 0', and 'aid = 1' implies 'obstacle = 0', and vice versa. Because of this, experimenting with these variables was not suitable for testing H1, since it would never be clear whether the changes in *involvement* and *distance* are caused by the increase in *good*, or by the decrease of *bad*, etc. Nevertheless, a number of experiments were performed with these variables, which confirmed that the behavior of that

part of the model globally corresponds to the theoretical model (Hoorn, 2008)⁴.

Discussion

In this paper, the theoretical model for involvement-distance trade-offs by (Hoorn 2008) has been translated into a simulation model in the LEADSTO language. Two main results were established. First, the model turned out to be adequate for simulating the dynamics of involvement-distance trade-offs and their influence on satisfaction. To model the trade-offs, the $(\gamma \cdot \max)$ version of Werners' (1998) fuzzy_AND operator seemed to provide more plausible results than the $(\gamma \cdot \min)$ version, since in ambiguous cases (where an agent experiences a more or less equal amount of involvement and distance simultaneously), it results in a relatively lower value for satisfaction than the $(\gamma \cdot \min)$ version. This is explained by the fact that the $(\gamma \cdot \max)$ version favors options in which involvement and distance differ much from each other. For example, it favors situations with $I=0.2$ and $D=0.8$ over situations with $I=D=0.5$, whereas for the $(\gamma \cdot \min)$ version this is the other way around. Second, and perhaps more surprisingly, it was found that positive features can increase the level of distance, and that negative features can increase involvement. This is explained by the fact that the factor levels do not directly influence involvement and distance, but only indirectly via the factors of similarity, relevance and valence. Although this finding may be counterintuitive, it corresponds to empirical evidence by (e.g., Van Vugt, Hoorn et al., 2006; Van Vugt, Konijn et al., 2006).

As mentioned above, our model was able to exhibit an increase in distance when the only change to the inputs of the model was an increase in a "positive" parameter. This success may seem arguable because the model is so complicated that almost any result is "possible." It might have been wiser to identify the simplest model that could exhibit this behavior, so as to identify the necessary/sufficient components that explain this phenomenon.

However true as this may seem theoretically, from an empirical point of view the model's complexity is warranted by years of experimentation (e.g., Konijn & Hoorn, 2005; Van Vugt et al., 2007). More important than complexity yet is the fact that the model excludes phenomena as well. Based on empirical evidence (ibid.), the model asserts that no more than 10 factors are needed to describe the full of human-robot interaction. These studies (ibid.) also show that realism is subordinate to ethical considerations, that no effects are established but through the mediation of relevance and valence, etc.

Yet, however nicely these empirical data were established, the theory as such still suffered from internal inconsistencies and logical blind spots. This is exactly what we repaired in the current paper. As such, simulations cannot count as a test on ecological validity but we did

⁴ See part C of the appendix.

show that we can simulate empirical results in a logically consistent way. In other words, model verification led to theory improvement.

To do so, we had to create a large number of different bias parameters that were set individually and parameters for setting the other individual characteristics, again underscoring the presumed over-complexity of the model. For one thing, the values of these parameters need to be settled empirically and because we could not do so right away, we set them to zero and one - to neutral that is - thereby reducing complexity again. But what we do have now, by making explicit hidden assumptions and creating internal consistency, is a framework to simulate the more complex affective behaviors and have a solid theoretical basis to do further empirical testing.

In future research, the model will be used to test other, more refined hypotheses. For example, it may be explored whether the use of bipolar variables instead of two unipolar variables (e.g., *ethics* instead of *good* and *bad*) provides different results. In addition, the process of extracting factor levels from features may be modeled in more detail, possibly taking more explicit goals of the robot into account. Another direction for future work is to combine the model with an existing computational model for emotion regulation (Bosse, Pontier, and Treur, 2007). Whereas the current model focuses on the *elicitation* of emotion, that model addresses the *regulation* of emotion. We expect that both models will smoothly fit together, since the satisfaction that is generated as output of the involvement-distance trade-off can almost directly be used as input to affective situation selection. In that case, current satisfaction is checked for a certain threshold and if it is too low, the robot will evaluate its expected satisfaction in alternative situations. Finally, in a later stage of the project, the model will be validated against empirical data of human affective trade-off processes. As soon as the model has been validated positively, we will start exploring the possibilities to apply it to real humans instead of agents, i.e., to develop a robot that can communicate affectively with humans.

Acknowledgements

The authors are grateful to the anonymous reviewers for their useful comments on an earlier version of this paper.

References

Appendix: see <http://www.cs.vu.nl/~tbosse/emobot>.
Bosse, T., Jonker, C.M., Meij, L. van der, & Treur, J. (2007). A Language and Environment for Analysis of Dynamics by Simulation. *International Journal of Artificial Intelligence Tools*, vol. 16, no. 3, pp. 435-464.
Bosse, T., Pontier, M., & Treur, J. (2007). A dynamical system modelling approach to Gross' model of emotion regulation. In: Lewis, R.L., Polk, T.A., Laird, J.E. (Eds.), *Proceedings of the 8th Int. Conference on Cognitive Modeling, ICCM'07*. Taylor and Francis, 2007, pp. 187-192.

Gratch, J. (2000). Modeling the Interplay Between Emotion and Decision-Making, *Proc. of the 9th Conference on Computer Generated Forces and Behavioral Representation*, 2000.
Gross, J.J. (1998). The Emerging Field of Emotion Regulation: An Integrative Review. *Review of General Psychology*, vol. 2, no. 3, pp. 271-299.
Gross, J.J. (2001). Emotion Regulation in Adulthood: Timing is Everything. *Current directions in psychological science*, vol. 10, no. 6, pp. 214-219.
Hooley, T., Hunking, B., Henry, M., & Inoue, A. (2004). Generation of Emotional Behavior for Non-Player Characters: Development of EmoBot for Quake II, *Proceedings of AAAI*, San Jose, CA, 2004.
Hoorn, J. F. (2008). *A Robot's Experience of its User: Theory*. In: Sloutsky, V., Love, B.C., and McRae, K. (eds.), *Proceedings of the 30th International Annual Conference of the Cognitive Science Society, CogSci'08* (this volume).
Konijn, E.A., & Bushman, B.J. (2007). World leaders as movie characters? Perceptions of G.W. Bush, T. Blair, O. Bin Laden, and S. Hussein at the eve of Gulf War II. *Media Psychology*. In press.
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.
Marsella, S., & Gratch, J. (2003). Modeling coping behavior in virtual humans: Don't worry, be happy. In *Proceedings of Second International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS'03*. ACM Press, pp. 313-320.
McCullagh, P., & Nelder, J.A. (1983). *Generalized Linear Models* (First ed.). London: Chapman and Hall.
Nelder, J. A., & Wedderburn, R. W. (1972). Generalized linear models. *J. R. Stat. Soc.*, A135: 370-384.
Van Vugt, H.C., Hoorn, J.F., Konijn, E.A., & De Bie Dimitriadou, A. (2006). Affective affordances: Improving interface character engagement through interaction. *International Journal of Human-Computer Studies* 64(9): 874-888. DOI: 10.1016/j.ijhcs.2006.04.008
Van Vugt, H.C., Konijn, E.A., Hoorn, J.F., Eliëns, A., & Keur, I. (2007). Realism is not all! User engagement with task-related interface characters. *Interacting with Computers* 19(2) 267-280.
Van Vugt, H.C., Konijn, E.A., Hoorn, J.F., & Veldhuis, J. (2006). Why fat interface characters are better e-health advisors. *Lecture Notes in Artificial Intelligence (LNAI)* 4133: 1-13. DOI 10.1007/11821830_1. Available at: <http://www.springerlink.com/content/g379221v1w65tu38/fulltext.pdf>
Werners, B.M. (1988). Aggregation models in mathematical programming. *Mitra*: 295-319.
Zimmermann, H.J. (1994). *Fuzzy Set Theory - and its Applications*. Boston, MA: Kluwer-Nijhoff.