

Humour, Relationship Maintenance and Personality Matching in automated dialogue: A controlled study

Marco De Boni *, Alannah Richardson, Robert Hurling

Unilever Corporate Research, Unilever Colworth, Sharnbrook, Bedford MK44 1LQ, UK

Received 27 February 2007; received in revised form 25 October 2007; accepted 3 November 2007

Available online 9 November 2007

Abstract

We built an automated dialogue system whose style of interaction can be varied along the three dimensions of Humour, Relationship Maintenance and Personality Matching. We then ran a longitudinal experiment which investigated manipulations of these three dimensions. We explored the interaction of these separate dimensions on user perception of the system using a controlled study design. We showed a strong positive effect for the use of Humour and Relationship Maintenance, while the use of Personality Matching raised a number of questions which need further investigation.

© 2007 Elsevier B.V. All rights reserved.

Keywords: Humour; Relationship Maintenance; Personality Matching; Automated dialogue

1. Introduction

Automated dialogue systems, systems which are capable of having an intelligent (within limits) text or speech-based conversation with a user, have been the object of research for a number of years (see for example [Androutsopoulos and Aretoulaki, 2003](#) for an overview of research in this area). From the point of view of technical implementation the field is relatively mature, with a number of “live” commercial systems and the availability of numerous tools to aid dialogue system implementation. There is a large body of literature on technical aspects of dialogue such as modelling dialogue flow, understanding user input or adapting dialogue strategies to the user and the formalization of pragmatic theories of dialogue.

Much less attention has been paid to the creation of the dialogue itself (as opposed to the design of the dialogue system): open issues include application and domain-independent content, both at the level of individual sentences

uttered by the system (should they be humorous? Matching the user’s language? Does it in fact matter at all?) and of general dialogue structure or strategy (how much social dialogue is necessary or acceptable from an automated system? Does it matter if systems don’t remember the content of previous interactions?).

Our work has explored in depth, through a longitudinal experiment, three important aspects of dialogue content: Relationship Maintenance (relationship maintenance over a number of sessions), Humour (fallibility, jokes) and Personality Matching (adapting the wording of the dialogue to the personality type of the user).

2. Prior work

Research that has tackled issues of dialogue content (as opposed to technical design) has included looking at the role of empathy in dialogue ([Liu and Picard, 2005](#)), the use of social dialogue (taken to mean chit-chat, empathetic utterances, [Bickmore and Picard, 2004](#)), trust ([Bickmore and Cassell, 2001](#)), emotions (e.g. designing an interaction to stimulate desired emotions in the user), personalisation (e.g. matching characteristics of the user) and narration

* Corresponding author. Tel.: +44 1234 248147; fax: +44 1234 248010.
E-mail addresses: macro.de-boni@unilever.com (M. De Boni), Bob.Hurling@Unilever.com (R. Hurling).

(e.g. linking a series of interactions by presenting previous content) (see [Stock, 1996](#)), familiarity (the way in which a relationship develops through reciprocal exchange of information), solidarity (a ‘like-mindedness’ or having similar dispositions, e.g. gender, psychological, political membership, etc.) and affect (the degree of liking or positive emotion generated during the interaction) (see [Svennevig, 1999](#)). Other researchers have examined adaptation, taking into account users’ relevant background knowledge to tailor system utterances, for example differentiating between novice and expert users (e.g. [Turunen et al., 2004](#)). In a different area researchers have examined co-operation in human–human and human–computer dialogue (e.g. [Bernsen et al., 1996](#)), looking at how to provide clear communication of what the system can do and how the user can interact with the system.

We found no agreement in the literature on the dimensions of dialogue content, nor any attempt to systematically categorise or relate in any way or even rank in order of importance, the aspects mentioned above. Consequently there is still much left unexplored regarding these aspects of dialogue content and there is a need for a strong evidence-based approach to fully understand what works and what does not in creating dialogues for automated dialogue.

Obviously, we could not tackle all of the possible aspects of dialogue content and therefore decided to concentrate on three aspects which intuitively appeared important, namely: relationship maintenance, humour and personality matching. Once we have understood these aspects we hope it will be clearer which dimensions to investigate next.

2.1. Relationship maintenance: continuity of dialogue over time

In maintaining a relationship over time a number of factors have been shown to be important. Aspects of conversation which underline a sense of persistence in a relationship and differentiate talk between strangers and acquaintances include references to mutual knowledge and talk about the past and future together ([Planalp and Benson, 1992](#); [Planalp, 1993](#)); greetings and farewells, as well as talk about the time spent apart have been shown to be important ([Gilbertson et al., 1998](#)). In looking at maintaining relationships over multiple interactions [Bickmore and Picard \(2005\)](#) have reported the positive effects of including relational behaviours (relational maintenance effects such as social chit-chat, empathetic feedback, meta-relational communication, humour) within computer–patient interactions: these led to significantly better working alliances and greater desire to continue working with the automated system. Looking at dialogue in a “real world” setting, researchers in psychotherapy and counselling have highlighted the role of Relationship Maintenance: a number of studies ([Bensing and Dronkers, 1992](#); [Hall et al., 1987](#); [Bertakis and Callahan, 1992](#); [Graugaard et al., 2005](#)) have shown how the communication between

physician and patients changes over time showing how the degree of task-focused communication significantly reduced between initial and return visits and that patients were more satisfied as consultations contained greater levels of socio-emotional communication (i.e. chatting) and less history taking.

In our work we have focused on understanding “Relationship Maintenance” in the sense of referring to and building upon content from previous sessions. This differs from other researchers, who for example have implemented (and evaluated) it together with several dialogue characteristics (empathy, chit-chat, humour, non-verbal behaviours, etc.) underneath an umbrella of ‘Relational Behaviours’.

2.2. Humour in dialogue: jokes and self-deprecation

A number of researchers have looked at the positive role of humour in human–computer communication as a relationship maintenance strategy ([Stafford and Canary, 1991](#); [McGuire, 1994](#); [Cole and Bradac, 1996](#); [Morkes et al., 1998](#)).

The role of self-deprecation as a type of humour has been examined by a number of studies. As with human–human interaction ([Amabile, 1983](#); [Amabile and Glazebrook, 1981](#); [Folkes and Sears, 1977](#); [Powers and Zuroff, 1988](#)) a computer is perceived as being more friendly when they praise others or criticize themselves than when they praise themselves or criticize others ([Nass et al., 1994](#)). [Stock \(1996\)](#) sees humour as playing an essential part of communication, acting both as a tension-releaser and generating positive associations with the dialogue system. Two routes for humour discussed are the direct telling of jokes (‘I have to exercise early in the morning before my brain figures out what I am doing’) and self-deprecation, (e.g. ‘Please could you word your answer differently – I’m a bit simple and can’t work out what you want!’). Humour has also been examined independently of self-deprecation. [Morkes et al. \(1998\)](#) have shown that canned (i.e. fixed and non-adaptive), text based humour “can be an inexpensive, high-impact method for increasing the likeability of an interface.” However using humour is recommended with a caveat of caution as it has also been shown to be perceived as often hard to understand, insulting and sometimes offensive.

The role of humour has also been recognized in practical dialogue settings: in Cognitive Therapy it is recognized by leading psychotherapists (e.g. [Beck, 1995](#)) and has been shown to be an important factor in the treatment of anxiety disorders ([Andrews et al., 2003](#)).

While the above research points to the fact that humour would probably have a positive impact on dialogue, we could not find any direct literature evaluating the impact of humorous elements within a dialogue system on user perception. We add to the literature by evaluating the user reaction to the same dialogue with or without the presence of humorous elements such as direct jokes and self-deprecating comments.

2.3. Personality Matching

Similarity in the communication style between interlocutors has been shown to have positive effects in a number of settings. Gill et al. (1999) have shown that emphasizing commonalities and de-emphasizing differences increases solidarity and rapport between communication partners, at least in a marital setting. Nass and Lee (2001) have shown that matching the personality of computer-generated speech (to appear either extrovert or introvert) to that of a recipient, generates the same consistency-attraction effects observed within therapist–client sessions (Herman, 1988). However participants in their experiments had limited exposure to, and were previously unfamiliar with, the speech system and it is not clear whether the influence of matched personality would continue over a more extended period. More importantly, Nass and Lee were not concerned with dialogue and instead explored the influence of personality matched speech within a primarily one-way communication format (e.g. book reviews on a web-site). Further work is consequently required to establish the impact of personality matching within automated dialogue. Again in the context of communication (but not dialogue) between a human and a computer, Moon and Nass (1996) and Moon (2002) have shown that customizing messages to match characteristics of the user's personality can make the message more persuasive. Participants in her study were classified as either Dominant or Submissive and then received messages written in either a matched or mis-matched style. Dominantly written messages were more persuasive for Dominant people whereas Submissive messages were more persuasive for Submissive people. Again, this study evaluated the persuasive influence of messages; further research is required to establish whether the same impact would be observed when matching parts of a dialogue to user's characteristics.

Matching health message style to the recipient's locus of control has been shown to make the message more persuasive (Williams-Piehota et al., 2004): people with an internal locus of control are more likely to believe that they are primarily in control of their health behavior, whereas those with an external locus tend to believe that powerful other people or forces are more responsible for their health status. For example, a health message emphasizing personal responsibility, rather than what others or 'the system' can offer, are more effective at persuading people with an internal locus of control, and vice versa.

We add to the literature by examining personality matching within a dialogue system, with the system presenting solutions written in either a dominant or submissive style. This differs from previous research in that the matched message is presented within a dialogue stream rather than a static message as implemented by Moon (2002).

3. Aims

We have concentrated our efforts on understanding the following three dimensions along which dialogues may be varied:

- Relationship Maintenance (building on previous interactions), representing the degree to which successive conversations between user and computer include and build upon content in previous interactions and where the impact of using and referring to content from previous interactions is evaluated, separate from the effect of empathy, chit-chat or non-verbal behaviours.
- Humour (the use of jokes and self-deprecation), where we examine the impact of humorous elements within a dialogue context.
- Personality Matching (interaction style similar to the users personality), where we assess the persuasive impact of matching the perceived personality of messages delivered within a dialogue to the personality of the user.

We built a dialogue system whose style of interaction can be varied along each of these three dimensions. We then ran an online longitudinal experiment which investigated manipulations of these three main dimensions for personalisation of the dialogue system. Rather than exploring the combined impact of several dialogue characteristics we explored the interaction of these separate factors on user perception of the system using a controlled study design. Users' experience of the system was then evaluated by a self-report framework, constructed in line with prominence-interpretation Theory (Fogg, 2003).

Obviously, the way in which we have chosen to influence relationship building, humour and personality matching is neither the only nor necessarily the best way to proceed, but we have chosen to focus on these limited elements to investigate if and how small(ish) changes along each dimension can have significant effects on users experience of the system.

4. Dialogue system overview

4.1. Overall design

The dialogue system under consideration was an automated exercise advisor service (Solution Advisor) which engaged users in a natural language conversation to help them overcome their barriers to exercise. This was done by eliciting the barriers that were stopping them from exercising and then proposing an appropriate solution to that particular barrier. The solutions were contained in a database compiled by domain experts and a combination of shallow parsing and statistical natural language processing methods were used to correlate the problems verbalized by the user with a particular solution. A full description of the system is beyond the scope of this paper, but in summary the components were:

(a) A Dialogue Manager which sets out the dialogue strategy, ensuring appropriate system responses to user input. The dialogue manager represents the dialogue as a state transition diagram where each state represents a possible system output and transitions are determined by user input.

(b) A User Model which contains information about the user such as their name, their personality profile, the problems they have told the system about and the solutions they were given

(c) A Natural Language Understanding component using shallow parsing (noun phrase chunking, part-of-speech tagging and limited dependency structure) and statistical natural language processing algorithms (k-nearest neighbor, trained on utterances compiled by a group of domain experts) to understand and categorize the user input (one or more English sentences); this component sent a representation of the knowledge gained from the user utterance to the Dialogue Manager

(d) A Natural Language Generation module which transforms the output of the Dialogue Manager (a particular state) into an appropriate English sentence.

4.2. Implementation of the desired features

Relationship Maintenance in the system was implemented:

- through the use of a user model and;
- by preserving a record of previous interactions.

This allowed the system to:

- Refer back to previous conversations, asking about the advice previously given and asking about the time between the previous interaction and the current one (following the ideas given in Gilbertson et al., 1998; Planalp and Benson, 1992; Planalp, 1993).
- Slowly build familiarity with the user by proceeding from a generic greeting introducing the system and its purpose in the first interaction to a gradually more friendly greeting with some elements of small talk in subsequent dialogues (following the results presented in Bensing and Dronkers, 1992; Hall et al., 1987; Bertakis and Callahan, 1992; Graugaard et al., 2005); this includes avoiding repetition, for example varying the wording in the type of greeting given to the user and in the advice given, but also covers small talk elements which change by becoming more personal over time.

One of the differences between the uses of Relationship Maintenance is shown by the following example of the login screen:

(Relationship Maintenance Off)

“Hello, this is the Solution Advisor, I can help you find a solution to overcome your exercise barriers. Just tell me what is preventing you from exercising more!”

(Relationship Maintenance On: the system recognizes the user and queries about the previous interaction)

“Good morning Michael.

In our previous interaction we came to the following solution to the barrier you had: [...] Did it prove effective for you?”

Humour was implemented through the use of self-contained jokes presented to the user at the end of each session and through self-deprecation during the session (the system admitted it might not be very good at recognizing answers). The utterance schemata provided a “slot” for both self-deprecation and humour, which could then be added or removed at will by the Generation module. We checked that the humour was appropriate by running a pilot study in which five people (who were not used in the subsequent main study) were asked to rank a list of jokes on how funny they were for use in a system giving advice on exercise behaviour. The eight most popular were chosen for the study.

The use of humour is demonstrated by the following examples:

(Using humour)

“Let’s check that I’ve understood (sometimes I can be a bit simple!): I would say that your barrier is Aches and Pains (e.g. You ache too much when you exercise). Would you agree?”

(Not using humour)

“Let’s check that I’ve understood: I would say that your barrier is Aches and Pains (e.g. You ache too much when you exercise). Would you agree?”

Personality Matching was implemented through appropriate schemata for transforming the solutions given by the system to the users’ problems into either a dominantly worded or a submissively worded message: following Moon (2002), “dominant” wording is given by the use of short and more blunt sentences and the use of strong words such as “will”, while “submissive” wording is given by longer, more gentle sentences and the use of softer forms such as “might want to”. The Generation module presented the user with the appropriate form depending on the context (e.g. previous utterances, whether the same utterance had been made in a previous interaction) and the given user model (dominant or submissive user).

The use of personality matching is demonstrated by the following example:

(Dominantly phrased solution)

“There are lots of other things you can do with your time, but it’s worth increasing your level of exercise because it brings so many benefits, including having more energy to enjoy the rest of your life. Raise the priority of exercise in your life over the next week and see the positive impact it has on your social life!”

(Submissively phrased solution)

“There could be lots of other things you can do with your time, but maybe it’s worth increasing your level of exercise because it could bring so many benefits, perhaps including having more energy to enjoy the rest of your life. Try raising the priority of exercise in your life over the next week and see if it impacts on your social life!”

5. Method

5.1. Experiment design

The object of the experiment was to investigate participant attitude to the conversational dialogues used in the Solution Advisor. Comparison between dialogues was conducted in a controlled and calibrated fashion recording user attitudes to each interaction. The experiment was a mixed, repeated measures design and investigated the effects of three main personalized dimensions of dialogue design.

- Relationship Maintenance (between groups)
- Personality Matching (dominance/submissiveness)
- Humour (jokes and fallibility)

6. Recruitment of participant sample

6.1. Criteria for inclusion

For this study participants were screened with inclusion criteria as follows:

- Internet savvy.* This ensured that participants had an expectation of what would be required of them in order to navigate and communicate with the simple web based user interface. We also wanted to ensure that those targeted in the study were participants who had easy access to the Internet in their daily lives as they would be required to log on to use the service 5 times over a duration of 2 weeks.
- Aged between 18 and 60 years of age.* 18 years of age restricted the recruitment of too many student participants and the age of 60 was advised by the surveying

agency as those who were panel participants with regular Internet experience.

- Be actively seeking to improve the amount of exercise they do.* This ensured a realistic scenario for the dialogue system.

The cohort was balanced for age, gender and presentation order of the experimental conditions between groups.

Participants were notified in advance of the experiment that they would be required to commit to interacting with the interface on 5 separate occasions (enrollment + 4 conversations) and that each interaction was set to take place 2 days following the last interaction.

6.2. Registration procedure

Immediately after passing through the inclusion criteria and prior to participants’ experience of the actual experiment conditions, they were asked to complete a registration procedure which incorporated a 60 item measurement of psychological androgyny which gives scores on dominance and submissiveness (Bem, 1974). This served to classify the participant into either a dominant or submissive group. Scores that were greater than or equal to zero were recorded as submissive.

Once registration was complete participants were randomly divided into two main effect groups where they were either exposed to a “continuous” sequence of interactions, where the system remembered previous interactions and acted accordingly, or a dialogue that did not remember previous interactions (Relationship Maintenance On/Off). Within these two main groups participants were then exposed to two randomized variables of (a) either having the dialogue matched to their personality type (along the dimensions of dominance and submissiveness) or mismatched (Matching On/Off) and (b) either being exposed to humour (jokes and fallibility) or not (Humour On/Off).

6.3. During the experiment

Participants were asked to evaluate the Solution Advisor after each interaction using a multiple item, seven-point, Likert-type attitude questionnaire (with positive and negative questions to avoid response acquiescence), the results of which were compared between sessions. After all four interactions had been completed an exit questionnaire was given to collect quantitative and qualitative participant attitudes on their interactions with the Solution Advisor.

The randomization of the states was determined as follows:

- Each user was required to complete four conversations with the dialogue system based on four set conversational topics (one for each of the four different states

being tested in the experiment). An example of conversational topic was: “Imagine that you have been trying to motivate yourself to start exercising but never seem to get round to it. See what the advisor recommends for this barrier”. The order of these interactions was decided from a precompiled shuffled state-sequence list. This ensured that there was equal coverage of all possible combinations of interaction sequences. The set conversation topics ensured that participants would (a) know what to communicate to the Solution Advisor and (b) allow for a relatively standardized experience across participants so as their experience of the barriers and their associated solutions were comparable (c) limit the utterance- we wanted to gauge the difference between key variants of the system. Therefore we had to eliminate the possibility of complex utterances which could have lead to ambiguity and erroneous classifications.

- Participants were asked to evaluate the Solution Advisor after each interaction with an attitude questionnaire, the results of which were compared between sessions.

6.4. The user interface

Before starting the experiment, we ensured the usability of the dialogue system through expert heuristic testing which provided feedback on factors such as the size of the text, navigation, information organization, control, page layout, appearance, readability, quality and enjoyment (Weir et al., 2006). The main experiment could therefore concentrate on the personalized dimensions of the dialogue interaction. For experimental control the interface was of a minimalist format with the webpage only showing the dialogue box from the Solution Advisor, the communication box for the user with a send button and the task present on the screen. No scrolling was involved during any task based interaction, to minimize variation between user experiences. This ensured that each task required the same attention to the Solution Advisor’s dialogue and a similar cognitive load being exerted on the user in the length and type of task that was required of them. This design ensured consistency and standardization while maximizing the user’s exposure to the alternative dialogues.

6.5. Instructions to users

One week prior to the study prospective participants were informed that the study aimed to examine an online advisor site designed to help people recognize their barriers and solutions to exercise. They were told that they would be required to use the site over a period of just over a week in order to register for the service and “talk” to the Solution Advisor over four conversations, with some set tasks.

Tasks were chosen to be typical of what people would expect to claim were reasonable barriers to their motivation to exercise or maintain a healthy level of physical activity; we also ensured the dialogue system was able to cope with the

task. The tasks were randomized between groups and participants to ensure that no order effects occurred.

An example of the interface is shown in Fig. 1. The task is shown in the left hand side. This was made persistent so as users did not forget the task they were being asked to perform. The text of the task was saved as an image so users were unable to cut and paste the task into their dialogue box. This ensured that users had to think and phrase the task in words that they had typed themselves.

6.6. Summary of experiment procedure

The procedure for each interaction with the Solution Advisor over the five sessions is outlined below:

- (1) Participants were asked to: Set up user names and passwords in order to access the interface over each session; Complete the dominance and submissiveness psychometric test.
- (2) Completion of first conversation task with Solution Advisor; Attitude Questionnaire.
- (3) Completion of second conversation task with Solution Advisor; Attitude Questionnaire.
- (4) Completion of third conversation task with Solution Advisor; Attitude Questionnaire.
- (5) Completion of fourth conversation task with Solution Advisor; Attitude Questionnaire; Overall Exit questionnaire.

There was a slight variation between the procedures of what was required from each group between the two main variables of Relationship Maintenance On and Relationship Maintenance Off. The conversational difference between the two group interactions is shown below:

Relationship maintenance on

- (1) User is asked what is stopping them from exercising; the system tries to classify the user’s answer as one of the pre-defined barrier categories; verifies this with the user;

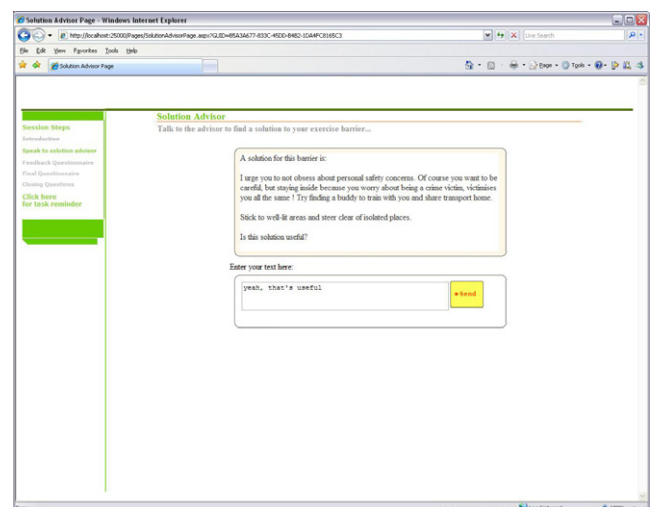


Fig. 1. An example of the user interface used in the experiment.

gives one barrier-specific solution; asks the user for its appropriateness/usefulness.

(2) This is a natural continuation of the first session whereby the system refers to the first interaction and asks how the user is progressing with it.

(3–5) Same principles of continuation used in the second session but encompass a more sociable welcome question.

Relationship maintenance off

(1) User is asked what is stopping them from exercising; the system tries to classify the user's answer as one of the pre-defined barrier categories; verifies this with the user; gives one barrier-specific solution; asks the user for its appropriateness/usefulness.

(2–5) Same as this first interaction.

6.7. Participant sample and descriptive statistics

A panel of 632 participants was recruited to assess the online conversational experience with the Solution Advisor. Of these 374 participants completed the necessary 5 sessions (Registration and 4 conversational tasks) in order for their inclusion in the data analysis of the service. Of these 189 participants were exposed to the Relationship Maintenance On condition (80 were male, 109 were female) while 185 participants (80 males and 105 females) were exposed to the Relationship Maintenance Off condition. Participants' ages ranged between 18 and 56 years of age, although the sample was not split to reflect different age categories. The sample of participants contained 214 females and 160 males. These were equally assigned into the main effect groups of Relationship Maintenance On and Relationship Maintenance Off. Participants' dominance and submissiveness was defined according to the Bem (1974) (BSRI); 162 participants were recorded as being dominant and 209 participants were recorded as being submissive.

6.8. Hypotheses

The hypothesis was that there would be a measurable difference in user attitude scores in response to the different conversational styles experienced in:

[H1:] The conversational style of dialogues matched to the user's dominant or submissive traits, versus dialogues that are deliberately mismatched to the user.

[H2:] Conversational style of dialogues that demonstrate temporal and relational Relationship Maintenance and dialogues that do not.

[H3:] Conversational style between a Solution Advisor that employs humour (jokes/fallibility) and one that does not.

6.9. Dependent variables

The dependent variables in this experiment were:

- Perceived attitudinal scores from a multiple attribute questionnaire (given after each exposure to the system, seven-point Likert-type scale) for both total scores and individual items.
- Qualitative and quantitative preferences between conditions defined from the exit questionnaire.

6.10. Independent variables

The independent variables in this experiment were:

- Experiment related: 2 main groups of subjects (Relationship Maintenance On and Relationship Maintenance Off).
- 4 types of dialogue, randomized presentation of dialogue and task, balanced for order effects.
- Participant related: 2 genders balanced, reflective of dominance/submissive characteristics, age group (2 groups balanced 18–35 and 36–56).

6.11. Questionnaire design

While there is a multitude of both objective and subjective dialogue evaluation metrics (Hajdinjak and Mihelc, 2006), these were not adequate for our purpose as we were examining particular aspects of a text-based dialogue system, not a spoken dialogue system and we were concerned with evaluating different content aspects of the dialogue, not optimizing different dialogue managers. In line with Prominence-Interpretation Theory (Fogg, 2003) we developed a self-report framework to measure both what the user had noticed at the end of each interaction with the dialogue system (Prominence) and the judgments users made about what they had noticed (Interpretation).

The attitude measurement used in this research adopts a seven-point Likert-type attitude scale in testing the user experience of the Solution Advisor. In order to define specific differences between experimental conditions individual statements from the questionnaire were analyzed separately. An inventory of multiple statements was chosen by a panel of evaluators (psychologists, computational linguists, usability experts) who assessed the questionnaire until inter-judge reliability was reached.

Individual statements were also analyzed separately to identify any specific issues that arose from the user sessions. The multiple items were consistent between all interactions allowing them to be compared both between groups and within variables. After experiencing both interfaces and completing the questionnaires participants were given a 16 item exit response questionnaire.

7. Results

7.1. R1: Participants experience of using the solutions advisor

Questions asked at the end of the experiment as part of the 16 item exit response questionnaire were analyzed using Logistic Regression (reported as χ^2 statistics below, degrees of freedom = 1 as there were only two possible responses) and ANOVA models to establish how the Relationship Maintenance On/Off manipulation was perceived by the participants (we considered using MANOVA but decided against this approach for our study, as this would only provide information on the average effect across all response items). Qualitative/open ended questions were assessed by independent judge assignment of responses to categories.

Participants were asked whether they had found the Solution Advisor similar or different to use throughout the trial. As hypothesized there was a significant difference, with those in the Relationship Maintenance Off group being more likely to report that the Solution Advisor felt similar to use (22% vs. 13% in the case of Relationship Maintenance Off; $\chi^2 = 5.00$, $p < 0.05$). There were no differences due to gender (male/female, $\chi^2 = 1.35$, $p = 0.25$) or personality (dominant/submissive, $\chi^2 = 1.50$, $p = 0.22$) of the participants.

All participants were asked whether they liked that the Solution Advisor asked if the previous session's advice had worked. The Relationship Maintenance Off group of course did not experience this in their interactions with the Solution Advisor. There was a highly significant difference between those in the Relationship Maintenance On group and those in the Relationship Maintenance Off group ($p < 0.0001$), with the Relationship Maintenance On group answering more positively (67% vs. 40%, $\chi^2 = 24.6$, $p < 0.001$), there was no effect due to gender ($\chi^2 = 0.95$, $p = 0.33$) or personality ($\chi^2 = 1.00$, $p = 0.32$) of participant.

Note that although the Relationship Maintenance Off group did not experience Solution Advisor dialogue that referred to previous interactions, a high proportion still stated that they liked being remembered between conversations.

All participants were asked whether they liked that the Solution Advisor had used their name during each interaction. Once again, those in the Relationship Maintenance Off group did not experience this Relationship Maintenance element. The difference between the two groups was highly significant, with 87% of Relationship Maintenance On participants noticing their name was used c/f 40% in the Relationship Maintenance Off group ($\chi^2 = 29.6$, $p < 0.001$), with the Relationship Maintenance On group also answering more positively. There was no effect of gender ($\chi^2 = 0.20$, $p = 0.65$) or personality ($\chi^2 = 0.39$, $p = 0.54$).

Participants were asked who they thought they were replying to during the study; a person or a computer pro-

gram. There was no significant difference between the Relationship Maintenance On and Off groups for the percentage of participants who thought there was a human operating the system (12% vs. 11%; $\chi^2 = 0.47$, $p = 0.49$).

7.1.1. Overall likes of the system

The Relationship Maintenance On and Off groups stated very similar likes in using the Solution Advisor. These fell into 6 main categories (% participants); Anonymity/Avoiding Embarrassment (28%), Convenient/Quick (25%), Easy to Use (19%), Getting Advice/New Ideas (16%), Fun/Novel (9%) and that using the Solution Advisor was like Talking to a Friend (4%).

Crucially, there was no significant difference in the proportion of participants reporting that the Solutions Advisor was "easy to learn to use" between the Relationship Maintenance On and Off groups (97 vs. 95%; $\chi^2 = 0.75$, $p = 0.39$). There was also no difference due to gender ($\chi^2 = 1.03$, $p = 0.31$) or personality ($\chi^2 = 0.53$, $p = 0.47$).

7.2. R2: Main hypotheses

An inventory of multiple statements assessed user attitude of the Solution Advisor. Individual statements were analyzed separately to evaluate whether the different experiment manipulations had influenced different items in the attitude questionnaire. The multiple items were consistent between all interactions, allowing them to be compared both between groups and within variables.

The questions were analyzed separately using a repeated measures analysis of variance model: the repeated measures adjusting for any temporal correlations between the 4 uses of the system by a panellist. The ANOVA modelled between-panellist effects of Relationship Maintenance Off/On, gender, panellist score on the dominance/submissiveness scale. Within-panellist effects included were Humour off/on, dominance/submissiveness matching off/on and order of use (i.e. 1–4).

7.2.1. H1: Personality Matching

Personality Matching as initially designed within our study (matching of the text used to describe the Exercise Barrier Solutions) had no significant influence on participants rating of the system [H1 not supported]. This was still the case when comparing participants from the top and bottom 20% of the Dominance-Submissiveness scale.

There was a significant difference between Dominant and Submissive participants' post session ratings of the system, regardless of whether the Exercise Barrier Solution had been matched or mis-matched. Note that there was no difference in dominant and submissive participants rating of how dominant the system was (4.87 vs. 4.84 respectively; $F_{1,1423} = 0.04$, $p = 0.83$).

The Dominant respondents found the Solution Advisor more 'supportive' (4.94 vs. 4.52; $F_{1,1423} = 12.21$, $p < 0.001$), more likely to 'change the quality of their life' (4.25 vs.

3.77; $F_{1,1423} = 11.37$, $p < 0.001$), said they were more likely to ‘try the advice’ (5.03 vs. 4.51; $F_{1,1423} = 12.54$, $p < 0.001$), found it more ‘helpful’ (4.74 vs. 4.28; $F_{1,1423} = 11.09$, $p < 0.001$), were more ‘satisfied’ (5.01 vs. 4.39, $p < 0.001$) and reported the system to be ‘better than they expected’ (4.92 vs. 4.38; $F_{1,1423} = 11.05$, $p < 0.001$). They were also more likely to say they would ‘recommend it to their friends’ (4.82 vs. 4.17; $F_{1,1423} = 18.13$, $p < 0.001$) and found the system more ‘persuasive’ (4.38 vs. 3.82; $F_{1,1423} = 15.72$, $p < 0.001$). Collectively these 8 items indicate that the dominant participants were more *Motivated* by the Solutions Advisor (Cronbach α for all 8 items = 0.95).

The Dominant respondents rated the Solutions Advisor as ‘easier to learn how to use’ (5.91 vs. 5.52; $F_{1,1423} = 14.87$, $p < 0.001$) and were less likely to agree that it ‘required a lot of concentration’ (2.50 vs. 2.82; $F_{1,1423} = 10.98$, $p < 0.001$). Dominant participants also rated the system as less ‘uncaring’ (2.60 vs. 2.97; $F_{1,1423} = 11.87$, $p < 0.001$) and were less ‘unsettled after using it’ (2.39 vs. 2.76; $F_{1,1423} = 11.07$, $p < 0.001$); indicating that overall they found it *less Cognitively & Emotionally Demanding* to use (Cronbach α for all 4 items = 0.78).

The Dominant respondents rated the Solutions Advisor as more ‘polite’ (5.53 vs. 5.23; $F_{1,1423} = 11.38$, $p < 0.001$), more ‘professional’ (5.29 vs. 4.89; $F_{1,1423} = 14.06$, $p < 0.001$), more ‘considerate’ (5.06 vs. 4.61; $F_{1,1423} = 19.42$, $p < 0.001$) and more ‘genuine’ (5.07 vs. 4.78; $F_{1,1423} = 6.69$, $p < 0.01$). Together these differences indicate that the dominant participants found the system more *Professional* (Cronbach α for all 4 items = 0.87).

Further analysis revealed that personality matching effects (of dominant or submissive participants with the overall system) were due to men and there was little difference between the responses of dominant and submissive women.

The Dominant participants also reported a much higher liking for the ‘Solution Advisor recognizing me’ (4.82 vs. 4.36; $F_{1,1423} = 13.59$, $p < 0.001$) and rated the system as more ‘approachable’ (5.04 vs. 4.71; $F_{1,1423} = 8.50$, $p < 0.01$), although these two items appear to indicate how *Socially Engaging* the system was, they had a relatively low Cronbach α of 0.59.

7.2.2. H2: Relationship Maintenance

In support of Hypothesis H2, the participants experiencing the Solution Advisor with Relationship Maintenance On rated the system as more ‘polite’ (5.45 vs. 5.30; $F_{1,1423} = 6.91$, $p < 0.01$), more ‘professional’ (5.16 vs. 5.01; $F_{1,1423} = 4.90$, $p < 0.05$), more ‘considerate’ (4.91 vs. 4.76; $F_{1,1423} = 4.94$, $p < 0.05$), and more ‘genuine’ (4.99 vs. 4.85; $F_{1,1423} = 4.72$, $p < 0.05$); indicating that they found the system more *Professional* than those with Relationship Maintenance Off. Those in the Relationship Maintenance On group also rated the system higher for the *Socially Engaging* items; ‘Solution Advisor recognising me’ (5.11 vs. 4.07; $F_{1,1423} = 186.21$, $p < 0.001$) and ‘approachable’ (4.95 vs. 4.80; $F_{1,1423} = 4.65$, $p < 0.05$).

7.2.3. H3: Humour

In support of Hypothesis H3, when humorous elements were present (in 2 out of the 4 sessions) the dialogue was perceived as having a more ‘appropriate level of humor’ (4.87 vs. 4.26; $F_{1,1423} = 71.89$, $p < 0.001$), ‘more enjoyable to use’ (5.13 vs. 4.93; $F_{1,1423} = 7.35$, $p < 0.01$), was rated lower for being ‘too impersonal’ (2.98 vs. 3.15; $F_{1,1423} = 4.70$, $p < 0.05$) and ‘too abrupt’ (2.87 vs. 3.05; $F_{1,1423} = 6.28$, $p < 0.01$). Together (Cronbach $\alpha = 0.76$) the responses to these items indicate that humor increases the overall level of *Emotive Bond* experienced during the dialogue.

7.3. R3: Dialogue progress

The order of the participants’ interactions were assigned from a precompiled shuffled state-sequence list, to ensure that there was equal coverage of all possible combinations of interaction sequences. ANOVA revealed 6 significant changes in participants’ perception of the Solution Advisor. The Solution Advisor was perceived as being significantly more ‘polite’ (averages for sessions in sequence order; 5.67, 5.40, 5.30, 5.14; $F_{3,1423} = 15.99$, $p < 0.001$) and ‘impersonal’ (3.31, 3.14, 2.90, 2.89; $F_{3,1423} = 6.88$, $p < 0.001$) during the first interactions, whilst using a more ‘appropriate level of humor’ (4.32, 4.52, 4.74, 4.68; $F_{3,1423} = 6.91$, $p < 0.001$), liked for ‘recognizing me’ (4.13, 4.61, 4.84, 4.77; $F_{3,1423} = 18.45$, $p < 0.001$) and feeling more ‘similar to me’ (3.71, 4.14, 4.26, 4.12; $F_{3,1423} = 9.08$, $p < 0.001$) increased in later interactions.

7.4. R4: Intention to use system again

The Relationship Maintenance On group reported a significantly greater preference to use the Solutions Advisor system again, with 53% stating they would use it on a weekly basis compared with only 32% for the Relationship Maintenance Off group ($\chi^2 = 6.9$, $p < 0.01$). There were no significant differences found between dominant and submissive participants ($\chi^2 = 0.19$, $p = 0.67$ or male and female participants ($\chi^2 = 0.08$, $p = 0.78$).

8. Discussion

The experiments we carried out have provided a number of interesting results underlining the importance of Relationship Maintenance and Humour and raised a number of questions about Personality Matching in natural language dialogue systems.

Of course, looking for correlations with a large set of items has the danger of giving spuriously significant results just through chance alone. However, our interest was the different patterns in significant items between the factors (Relationship Maintenance, Personality Matching, Humour) rather than simply finding any significant difference. It is however necessary to follow up with further studies that test the reliability of our findings.

Overall, the average differences in users ratings of the system variants are small (typically less than one unit on a seven-point scale). However, it should be noted that in this study, we were interested in the impact of relatively small changes to the system, so for example the manipulation of humour was represented by the inclusion of a joke in only one or two sentences of a dialogue interaction that involved many sentences and multiple screens. Therefore, we did not expect the small system changes we had made to lead to large changes in user perception. The degree of impact could however be explored within the same experimental paradigm; for example the proportion and/or intensity of humour (or continuity or personality matching) could be increased and the subsequent change in user perception monitored. Also, the means represent the average for the user population and so are influenced both by variation between each individual's scoring style as well as their different reactions to the system. The scores also represent ratings of each individual system just after they have been experienced, which increases the validity of the rating (by maximizing its proximity to the experience) but this also lead to most variants of the system being rated before having experienced all the other variants.

A significant result was derived from the use of Relationship Maintenance. Relationship Maintenance has been considered an important part of dialogue, as demonstrated by the work of Bickmore et al. (2005) on automated dialogue and the work on physician-patient communication by, amongst others, Bensing and Dronkers (1992); Hall et al. (1987); Bertakis and Callahan (1992); Graugaard et al. (2005). Our work has shown that this intuition is correct: The system which built a relationship over time with the users by processing and elaborating information from previous sessions and relating the current situation with the content of previous sessions was perceived as more professional, being more polite, considerate and genuine. Designers want their system to be used again and again and want to encourage users to re-engage with the system; by building systems which engage with the user over multiple sessions, remembering them and building some form of relationship over time we have shown that systems demonstrate a much higher sense of professionalism, which in turn should encourage users to make use of the system again in future.

Humour, using jokes and self-deprecation, also had a positive outcome, leading to a greater emotional bond with the system. This is consistent with the work cited above on computer interfaces by, for example, Morkes et al. (1998); Stock (1996), as well as the work on human–human interaction by, amongst others, Stafford and Canary (1991); McGuire (1994); Cole and Bradac (1996); Morkes et al. (1998). It is interesting to note that the effects of Relationship Maintenance, indicating that systems may not rely solely on either Humour or Relationship Maintenance and need to implement both for best results.

In this particular implementation the humorous elements were canned (i.e. fixed, predefined) jokes which were not adapted either to the context of the dialogue (jokes did not relate directly to the particular dialogue content but were generically about health) or the particular user's situation (they did not adapt to the language or to any other characteristics of the user); the self-deprecating elements were comments which were added to the generated system output. Presumably adapting the humorous and self-deprecating elements to the user and dialogue context would have an even greater effect, but the fact that such a simple template-based approach works shows the power of Humour.

There was no effect however for Personality Matching. This seems to contradict the results of Moon (2002) who reported on the persuasiveness of messages tailored to the dominant or submissive personality of users. As the system as a whole was perceived as dominant, it may be the case that the variations in phrasing of the solution, which should have given part of the system submissive as well dominant characteristics, were either too subtle or constituted too small a part of overall dialogue to have an effect. On the other hand Moon (2002) worked on prose, not dialogue, and it may be the case that this type of matching works at the message level, but for whole dialogues.

It is difficult to say with certainty why dominant people were overall more positive about the system: this could be due to the unintentional matching effect between the system (perceived overall as being dominant) and the dominant users; but an equally plausible explanation could be that dominant people simply like technology more than others and therefore would be more likely to rate a system favourably. We will need to carry out further research in this area to be able to draw firm conclusions; in particular, we will need to build a system with overall dominant or submissive characteristics as opposed to the subtle differences which we implemented in the current version.

9. Conclusion

We built a dialogue system whose style of interaction can be varied along the three dimensions of Humour, Relationship Maintenance and Personality Matching and then ran a longitudinal experiment which investigated manipulations of these three dimensions using a controlled study design. We showed that using Relationship Maintenance and Humour has a positive result on the perception of the dialogue system. Having established that these elements are important to dialogue, further work will examine these elements in more depth. Questions we want to answer include: what is the relationship between humour and system performance? Although overall humour is a positive addition to dialogue, are there times when its use is inappropriate? Other questions arise regarding relationship maintenance: how can we maximise continuity between sessions? How much and what type of small talk should we introduce in a system?

The results for Personality Matching on the other hand were inconclusive. This could be for a number of different reasons: perhaps matching does not work if it is only carried out on a limited part of the dialogue; perhaps the effects of matching dimensions such as dominance and submissiveness are confounded by other important elements such as politeness (a dominant, forceful interaction might be perceived as less polite); perhaps matching works for narrative but cannot be transferred in this way for dialogue. Further work will be needed to examine personality matching in more detail in order to draw some more solid conclusions.

Appendix A. Questionnaire

I thought the Solution Advisor was enjoyable to use.
 I thought the Solution Advisor was very polite.
 I felt that the Solution Advisor was too impersonal.
 I found the Solution Advisor supportive.
 Using the Solution Advisor would change the quality of my life.
 I thought the Solution Advisor was too abrupt.
 I would like to use the Solution Advisor again.
 I thought the Solution Advisor was reliable.
 I thought that the Solution Advisor was frustrating to use.
 I felt in control of the conversation with the Solution Advisor.
 I thought that the Solution Advisor was incompetent.
 It was easy to learn how to use the Solution Advisor.
 The Solution Advisor has influenced me to change the beliefs I had before.
 The Solution Advisor was not easy to understand.
 I felt that the Solution Advisor had a dominant attitude.
 The Solution Advisor used an appropriate level of humour.
 Using the Solution Advisor required a lot of concentration.
 The Solution Advisor was too long winded.
 I will try the advice that the Solution Advisor recommended.
 The Solution Advisor talked to me in a professional manner.
 The Solution Advisor was very informative.
 The Solution Advisor provided reassuring advice.
 The Solution Advisor's responses were irrelevant to what I had entered.
 I thought the Solution Advisor was sincere.
 I thought the Solution Advisor adapted to my needs.
 I thought that the Solution Advisor was honest.
 I found the Solutions Advisor's questions too intrusive.
 I found the Solution Advisor helpful.
 I thought the Solution Advisor was dependable.
 I thought the Solution Advisor was considerate.
 I thought the Solution Advisor was uncaring.
 I felt that the Solutions Advisor was approachable.

The advice offered by the Solution Advisor seemed genuine.
 I felt that the Solution Advisor was trustworthy.
 I thought that the Solution Advisor was patient.
 Conversation with the Solution Advisor was easy to follow.
 I felt unsettled after using the Solution Advisor.
 It was easy to engage with the Solution Advisor.
 Overall I was very satisfied with my experience of the Solution Advisor.
 Using the Solution Advisor for advice was better than I expected.
 I felt the Solution Advisor had a submissive attitude.
 I think that the Solution Advisor is a motivating influence to exercise more.
 I found the conversation with the Solution Advisor uncomfortable.
 I liked that the Solution Advisor recognised me from the last conversation we had.
 I would prefer to talk to a human about my exercise barriers rather than the Solution Advisor.
 Using the Solution Advisor has made me confident that I can change from my existing bad habits.
 I would recommend using the Solution Advisor to my friends.
 The Solution Advisor seemed similar to me.
 I found the Solution Advisor persuasive.

Appendix B. Exit questionnaire

What did you like about interacting with an online service such as this?
 What did you dislike about interacting with an online service such as this?
 Did you find learning how to use the Solutions Advisor easy or difficult?
 Did the four Solution Advisors feel similar to use or different to use?
 At the beginning of each session the Solution Advisor asked you if the previous session's advice had worked for you. How did this make you feel?
 The Solution Advisor used your name in the conversations. How did this make you feel towards the Solutions Advisor?
 What do you see as the benefits of discussing your exercise barriers with a system such as this?
 What do you see as the drawbacks of discussing your exercise barriers with a system such as this?
 Is there any way that the system could be improved?
 Who do you think you were replying to?
 Would you use the Solutions Advisor in real life?
 Please write down an exercise barrier that you would personally like to discuss with the Solution Advisor
 How often do you think you would use a service such as this?

How much would you be prepared to pay to use a service such as this per month?

Please write down any words that spring to mind to describe the Solution Advisor.

How would you score the Solution Advisor on the scale below.

References

- Amabile, T.M., 1983. Brilliant but cruel: Perceptions of negative evaluators. *Journal of Personality and Social Psychology* 19, 146–156.
- Amabile, T.M., Glazebrook, A.H., 1981. A negativity bias in interpersonal evaluation. *Journal of Personality and Social Psychology* 18, 1–22.
- Andrews, G., Creamer, M., Crino, R., Hunt, C., Lampe, L., Page, A., 2003. *The Treatment of Anxiety Disorders: Clinician Guides and Patient Manuals*, second ed. Cambridge University Press, Cambridge.
- Androutsopoulos, I., Aretoulaki, M., 2003. Natural Language Interaction. In: Mitkov, R. (Ed.), *Handbook of Computational Linguistics*. Oxford University Press, pp. 629–649 (Chapter 35).
- Beck, J., 1995. *Cognitive Therapy: Basics and Beyond*. Guildford Press, New York.
- Bem, S.L., 1974. The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology* 42, 155–162.
- Bensing, J.M., Dronkers, J., 1992. Instrumental and affective aspects of physician behavior. *Medical Care* 30, 283–298.
- Bernsen, N.O., Dybkjær, H., Dybkjær, L., 1996. Cooperativity in human-machine and human-human spoken dialogue. *Discourse Processes* 21 (2), 213–236.
- Bertakis, K.D., Callahan, E.J., 1992. A comparison of initial and established patient encounters using the Davis observation code. *Farm Medicine* 24, 307–311.
- Bickmore, T., Picard, R.W., 2004. Establishing and Maintaining Long-Term Human-Computer Relationships. *Transactions on Computer-Human Interaction* 12 (2), 293–327.
- Bickmore, T., Cassell, J., 2001. Relational Agents: a model and implementation of building user trust. *Proceedings of CHI-2001*. ACM Press, USA, pp. 396–403.
- Bickmore, T.W., Picard, R.W., 2005. Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction* 12, 293–327.
- Bickmore, T.W., Gruber, A., Picard, R., 2005. Establishing the computer-patient alliance in automated health behavior change interventions. *Patient Education and Counseling* 59, 21–30.
- Cole, T., Bradac, J., 1996. A Lay Theory of relational satisfaction with best friends. *Journal of Social and Personal Relationships* 13 (1), 57–83.
- Fogg, B.J., 2003. Prominence-interpretation theory: explaining how people assess credibility online. *CHI 2003*, Fort Lauderdale, Florida, pp. 722–723.
- Folkes, V.S., Sears, D.O., 1977. Does everybody like a liker? *Journal of Experimental Social Psychology* 13, 505–519.
- Gilbertson, J., Dindia, K., Allen, M., 1998. Relational Continuity Constructional Units and the Maintenance of Relationships. *Journal of Social and Personal Relationships* 15 (6), 774–790.
- Gill, D., Christensen, A., Fincham, F., 1999. Predicting marital satisfaction from behavior: do all roads really lead to Rome? *Personal Relationships* 6, 369–387.
- Graugaard, P.K., Holgersen, K., Eide, H., Finset, A., 2005. Changes in physician-patient communication from initial to return visits: a prospective study in haematology outpatient clinic. *Patient Education and Counseling* 57, 22–29.
- Hajdinjak, M., Mihelic, F., 2006. The PARADISE evaluation framework: issues and findings. *Computational Linguistics* 32 (2).
- Hall, J.A., Roter, D.L., Katz, N.R., 1987. Task versus socio-emotional behaviors in physicians. *Medical Care* 25, 399–412.
- Herman, S.M., 1988. The relationship between therapist-client modality similarity and psychotherapy outcome. *The Journal of Psychotherapy Practice and Research* 7, 56–64.
- Liu, K., Picard, R.W., 2005. Embedded empathy in continuous, interactive health assessment. *CHI Workshop on HCI Challenges in Health Assessment*, Portland, Oregon.
- McGuire, A., 1994. Helping behaviors in the natural environment: dimensions and correlates of helping. *Personality and Social Psychology Bulletin* 20 (1), 45–56.
- Moon, Y., Nass, C., 1996. How ‘real’ are computer personalities? Psychological responses to personality types in human-computer interaction. *Communication Research* 23, 651–674.
- Moon, Y., 2002. Personalization and personality: some effects of customizing message style based on consumer personality. *Journal of Consumer Psychology* 12 (4), 313–326.
- Morkes, J., Kernal, H.K., Nass, C., 1998. Humor in task-orientated computer-mediated communication and human computer interaction. *Conference on Human Factors in Computing Systems*, Los Angeles.
- Nass, C., Lee, M.K., 2001. Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, and consistency-attraction. *Journal of Experimental Psychology: Applied* 7, 171–181.
- Nass, C., Steuer, J., Henriksen, L., Dryer, D.C., 1994. Machines, social attributions, and ethopoeia: performance assessments of computers subsequent to “self-” or “other-” evaluations. *International Journal of Human-Computer Studies* 40, 543–559.
- Planalp, S., Benson, A., 1992. Friends’ and acquaintances’ conversations I: perceived differences. *Journal of Social and Personal Relationships* 9, 483–506.
- Planalp, S., 1993. Friends’ and acquaintances’ conversations II: coded differences. *Journal of Social and Personal Relationships* 10, 339–354.
- Powers, T.A., Zuroff, D.C., 1988. Interpersonal consequences of overt self-criticism: a comparison of neutral and self-enhancing presentations of self. *Journal of Personality and Social Psychology* 54 (6), 1054–1062.
- Stafford, L., Canary, D., 1991. Maintenance Strategies and Romantic Relationship Type, Gender and Relational Characteristics. *Journal of Social and Personal Relationships* 8, 217–242.
- Stock, O., 1996. Password Swordfish: Verbal humor in the interface. In: *Proceedings of the International Workshop on Computational Humor*, TWLT – 12, Enschede.
- Svennevig, J., 1999. *Getting Acquainted in Conversation. A Study of Initial Interactions*. John Benjamins, Amsterdam.
- Turunen, M., Salonen, E.-P., Hartikainen, M., Hakulinen, J., Black, W., Ramsay, A., Funk, A., Conroy, A., Thompson, P., Stairmand, M., Jokinen, K., Rissanen, J., Kanto, K., Kerminen, A., Gamback, B., Cheadle, M., Olsson, F., Sahlgren, M., 2004. AthosMail – a multilingual adaptive spoken dialogue system for E-mail domain. In: *Proceedings of Workshop on Robust and Adaptive Information Processing for Mobile Speech Interfaces*, Geneva, pp. 77–86.
- Weir, C., Anderson, J., Jack, M., 2006. On the role of metaphor and language in design of third party payments in eBanking: usability and quality. *International Journal of Human-Computer Studies* 64 (8), 770–784.
- Williams-Piehot, P., Schneider, T.R., Pizarro, J., Mowad, L., Salovey, P., 2004. Matching health messages to health locus of control beliefs for promoting mammography utilization. *Psychology and Health* 19, 407–423.