Available online at www.sciencedirect.com
SCIENCE DIRECT
Interacting with Computers 17 (2005) 522–541
Interacting with Computers
www.elsevier.com/locate/intcom

Assessing the effects of building social intelligence in a robotic interface for the home

Boris de Ruyter[b,*], Privender Saini[a], Panos Markopoulos[a], Albert van Breemen[b]

[a]User System Interaction Programme, Eindhoven University of Technology, Den Dolech 2, 5600 MB Eindhoven, The Netherlands
[b]Philips Research, Media Interaction, Prof. Holstlaan 4 (WY 2.01), 5656 AA Eindhoven, The Netherlands

Available online 13 May 2005

Abstract

This paper reports an exploration of the concept of social intelligence in the context of designing home dialogue systems for an Ambient Intelligence home. It describes a Wizard of Oz experiment involving a robotic interface capable of simulating several human social behaviours. Our results show that endowing a home dialogue system with some social intelligence will: (a) create a positive bias in the user's perception of technology in the home environment, (b) enhance user acceptance for the home dialogue system, and (c) trigger social behaviours by the user in relation to the home dialogue system.

Keywords: Ambient Intelligence; Social intelligence; Human-like

1. Introduction

The vision for Ambient Intelligence is that humans will be surrounded by embedded technology in the home (Aarts et al., 2001). There will be an ever-increasing number of devices that we will need to know how to operate in our daily lives. This complexity implies that the human interaction with this technology will bring about a shift in the way we communicate with it. There are two prevailing views on this communication (Markopoulos, 2004). One in which communication is through a multitude of task-specific

* Corresponding author.
E-mail addresses: boris.de.ruyter@philips.com (B. de Ruyter), p.k.saini@tue.nl (P. Saini), p.markopoulos@tue.nl (P. Markopoulos), albert.van.breemen@philips.com (A. van Breemen).

0953-5438/$ - see front matter
doi:10.1016/j.intcom.2005.03.003

information appliances (Norman, 1998). The other view is that a centralized user interface anticipates users' needs through adaptivity and intelligence (Pentland, 2000). In the home domain, *home dialogue systems* are expected to fulfil this latter role, acting as intermediaries between systems and services incorporated in the home and for the people in them. There are several issues regarding the acceptance of home dialogue systems by users. It is clear that users need to like and trust such systems in order for them to deliver their intended benefits. This research explores the hypothesis that appropriate design of the social behaviour of these agents could lead to improved acceptance, not just of the home dialogue system itself but also of the embedded technology it helps users access.

Since Reeves and Nass (1996) published their work on the *media interaction* a lot of research has explored the notion that people react to media as if they were social actors. There is a growing interest in on-screen agents and robotic characters that are endowed with human-like emotions and behaviour. Some such prototypes have been made for the purpose of cultivating personal relationships (like friendship) with the user (NEC, 2004; Omron, 2004). They come with their own personalities, unique traits and functionalities. Others are found in a public domain, guiding users through their realm (Bickmore and Cassell, 2001; Heylen et al., 2004), providing information on products, offering services like online ticket sales. Others, like Breazeal et al. (submitted for publication) and Breazeal and Scassellati (2000), represent attempts by researchers to implement various 'human-like' characteristics into a robot (dialogue systems, facial expression recognition and presentation, gesture generation and recognition, humour perception and production, etc.). The behaviours that these social/emotional characters (screen agents and robots) display range from facial animation (including eye-gaze and head movements) (Bickmore and Cassell, 2001; Heylen et al., 2004; Bickmore and Picard, 2004; Sidner et al., 2004; Thórisson, 1997) to limb and hand gestures (Beskow and McGlashan, 1997; Bickmore and Cassell, 2001; Bickmore and Picard, 2004; Thórisson, 1997); full-body posture adjustments (Bickmore and Cassell, 2001; Bickmore and Picard, 2004; Breazeal et al., submitted for publication) and locomotion (Bruce et al., 2002; Hűttenrauch and Eklundh, 2002). The standard research paradigm has been to compare characters with these expressive behaviours with characters lacking these. In the majority of cases the superiority of the expressive character over its neutral counterpart is demonstrated in terms of characteristics such as likeability and trust (Bickmore and Cassell, 2001; Bickmore and Picard, 2004; Thórisson, 1997), naturalness of interaction (Bickmore and Cassell, 2001), satisfaction with interaction (Bickmore and Cassell, 2001; Heylen et al., 2004), ease of use (Heylen et al., 2004), efficiency of task completion (Heylen et al., 2004), closeness to human characteristics (Thórisson, 1997), appropriateness of movements (Thórisson, 1997; Sidner et al., 2004; Heylen et al., 2004), joint attention (Imai et al., 2001), attention getting and cooperation elicitation from strangers (Bruce et al., 2002) and invested effort in interaction task (Bartneck, 2003).

Common to the above works is that test-users reacted positively towards more expressive and human-like characteristics. Nevertheless, such research is limited to displaying relatively low-level behaviours and assessing fairly direct effects of these behaviours. The present work examines social intelligence as a higher-level construct to explain why, in some cases, a more human-like system may be preferred over others.

2. Social intelligence

When we consider human–human social interactions, we see that there are several characteristics that make certain individuals stand out and more liked by others, or which convey an air of trustworthiness, competence and dependability (Ford and Tisak, 1983; Keating, 1978; Sternberg and Smith, 1985). This list is large and includes attributes like ‘being nice and pleasant to interact with’, ‘being on time for appointments’, ‘thinking and speaking before doing’ and ‘being sensitive to other people’s needs and desires’. These attributes and many others are part of the concept of social intelligence; they can be seen as a continuum with the socially intelligent person falling in the high range. In its broadest definitions social intelligence is ‘…a person’s ability to get along with people in general, social technique or ease in society, knowledge of social matters, susceptibility to stimuli from other members of a group, as well as insight into the temporary moods of underlying personality traits of strangers’ (Vernon, 1933). So the socially intelligent person has a better than average ability to judge other peoples feelings, thoughts, attitudes and opinions, intentions, or the psychological traits that may determine their behaviour. This judgement creates expectations on the observer’s part about the likely behaviour of the observed person. This in turn leads to adjustments of one’s own behaviour accordingly and appropriately. However, that appropriateness can only be judged when the social context is taken into account. In this sense, social intelligence is not merely something that goes on between two people in isolation, but contextual factors also come into play.

Clearly, a person needs at least some degree of social intelligence to make his or her way in the world. This applies equally to socially complex situations (like business meetings) as to seemingly simple chores like grocery shopping.

A considerable amount of research literature has been published in the field of psychology on social intelligence. This literature, however, is rather vague on the exact characteristics and behaviours that constitute social intelligence. While it would be feasible, for example, to draw up a list of behaviour characteristics that would indicate mathematical intelligence—where the situations in which it is manifested are more confined, and thus relatively easily defined—this would be a daunting task when it comes to social intelligence. The situations imaginable are much more varied and are subject to constant change as interaction takes place. The manifestation of social intelligence is therefore almost entirely dependent on the context, making it difficult to provide an exhaustive list of characteristics of social intelligence. As a consequence there are many ways in which we can make an interactive system exhibit social intelligence. The following sections describe how we designed and implemented a set of behaviours to make a robot be perceived as socially intelligent.

The purpose of the study is to examine the effects of social intelligence in a robotic home dialogue system. Although there has recently been significant interest in social intelligence in the domain of computational and robotic characters, the benefits that such intelligence might bring have not been demonstrated or clearly identified. This study aims to address this apparent shortcoming. Rather than focusing on the direct effect of one or two behaviours, we took a broader approach: a number of aspects of social intelligence were implemented in a robotic character. An experiment was conducted to examine the effects of social intelligence; more specifically we addressed the following research questions:

- Will the level of social intelligence implemented in the home dialogue system be perceived?
- What is the effect of bringing the concept of social intelligence into a home dialogue system on the perception of the quality of the interactive systems (other than the home dialogue system) in the environment?
- Will acceptance of home dialogue systems increase if the concept of social intelligence is implemented in these systems?

The home dialogue system used to answer these questions is described in the next section. In the sections which follow we also describe the development of the Social Behaviors Questionnaire. After this an experiment is described addressing the research questions raised above. The results of this experiment are discussed, leading to some conclusions in the final section.

3. Robots versus on screen characters

For this study we used a robotic character to display social behaviours. Although there are obvious benefits of screen characters in terms of implementation facility, robots have properties of their own that make them compelling as candidates for a home dialogue system. One of the most obvious differences is the physical embodiment of a robot in the world. Bartneck (2002) found that there was no difference in how his test participants enjoyed interaction with a robot and a similar onscreen character, but there was a social facilitation effect. He found that participants would put more effort into a negotiation task with a robotic character than with a screen character. Another benefit of robots in terms of the social side of the interaction with their users is locomotion (walking or rolling around). Instead of the user approaching the interface for contact, the interface can approach the user. Taking this a step further, the user does not have to focus on the robot throughout the interaction. S/He can walk around and the robot can follow or turn its head as an indication that it is still 'listening' or 'attending' to the user. This turning of the head to track the user can also be used to look at other objects for communication purposes. Gestures can be performed that will make communication less ambiguous than communication with a screen character. Another option that flows from the locomotion benefit is the transport and manipulation of small objects for the convenience of the user. Other interaction techniques also come into play that are likely to make the interaction more natural than with screen characters. The robot can be touched. And the different touches can signify different commands, requests etc. In short, robots can open up a large and interesting design space to facilitate a more natural form of interaction with their users. It should be noted, however, that even though locomotion provides an important benefit over screen characters, we did not use this property in the current studies. The main reason for this is because our goal was to create a socially coherent character that would be present peripherally even when the participant's attention was not focused on the robot. For our purpose, locomotion was not a requirement.

4. The iCat

The home dialogue system used in our study takes the form of an 'interactive Cat', or just iCat. The iCat is a research platform for studying social robotic user-interfaces. It has a cat-like appearance because of the acceptance of cats in domestic environments.

The iCat is a 38 cm tall user-interface robot that lacks mobility facilities. This way, we can solely focus on the robot-human interaction aspects during the development of this platform. The robot's head is equipped with 13 standard R/C servos that control different parts of the face, such as the eyebrows, eyes, eyelids, mouth and head position. With this set-up we can generate many different facial expressions that are needed to create an expressive character. Fig. 1 illustrates facial expressions for the basic emotions happiness, surprise, fear, sadness, disgust, anger and neutral. A camera is installed in the iCat's head that is used for different computer vision capabilities, such as recognizing objects and faces. iCat's foot contains two microphones to record the sounds it hears, perform speech recognition and to determine the direction of the sound source. Also, a speaker is installed

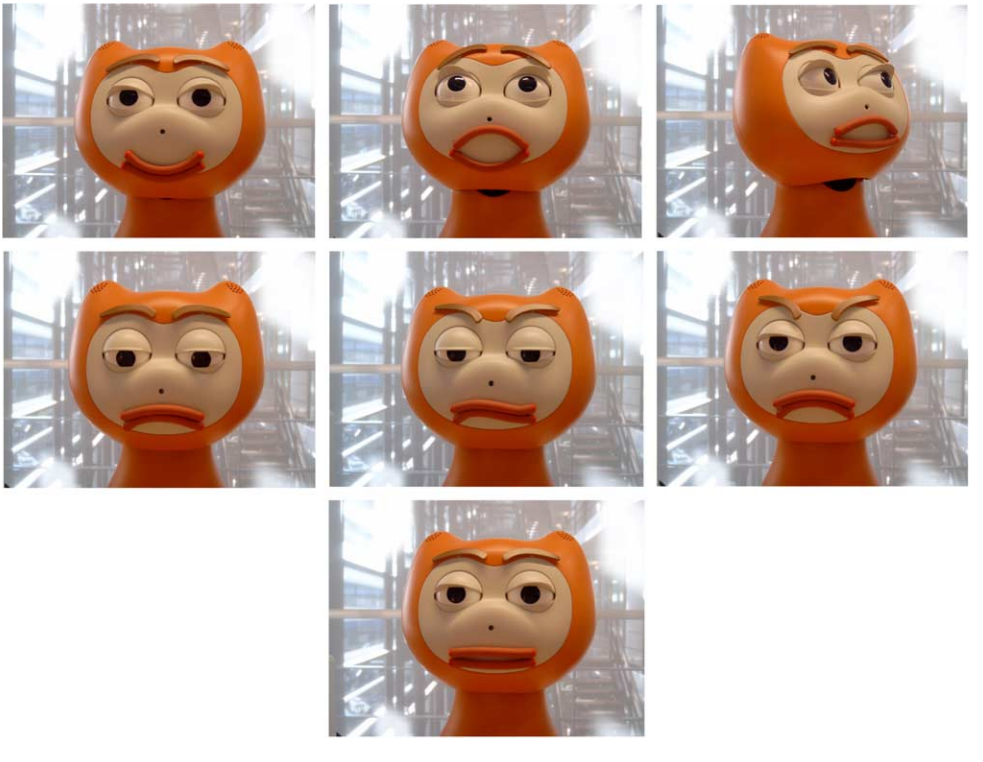

Fig. 1. The iCat's facial expressions. From left to right: happiness, surprise, fear, sadness, disgust, anger, neutral.

to play sounds (WAV and MIDI files) and to generate speech. Furthermore, iCat is connected to a home network to control in-home devices (e.g. lights, VCR, TV, radio) and to obtain information from the Internet. Finally, touch sensors and multi-colour LEDs are installed in the feet and ears to sense when the user touches the robot and to communicate further information encoded in coloured light. For instance, the operation mode of the iCat (e.g. sleeping, awake, busy, listening, etc.) is encoded in the colour of the LEDs in the ears.

Traditionally robots are controlled using feedback control techniques, which result in rather machine-like movements of the robot's body parts (e.g. constant velocities). This makes the perception of a robot more like a machine than an intelligent and socially conversant agent. In order to create a socially communicative robot, the robot should be controlled in way which makes its behaviour credible.

The notion of credibility means that the actions of the robot should be apparent and understandable to a user. Credibility can be achieved by applying animation principles to the behaviour generation module of the robot. Van Breemen (2004) has shown that traditional animation principles can indeed be applied to robots to improve their credibility. For the iCat, a dedicated robot animation engine was developed to control all of the iCat movements and to generate socially intelligent behaviours. A dedicated tool, called the Robot Animation Editor, was developed to graphically design believable robot animations, such as eye-blinking, facial expressions, head movements, etc.

5. Measuring perceived social intelligence: the social behaviors questionnaire (SBQ)

In order to verify whether one system or the other conveys the impression of social intelligence to its users, a measurement instrument was needed. In psychology, there have been different approaches to measuring the social intelligence of individuals. For example, photographs and stories have been used to see how well people perceive non-verbal cues and how they respond to moral dilemmas. But by far the most commonly used measure is the questionnaire. Although there are a number of good questionnaires assessing social intelligence (Moos et al., 1927; Gough, 1968; Strang et al., 1942; Banham, 1968), they seem to primarily be made up of questions one answers about oneself (self-reports). The current study calls for a measurement of the perception of the degree of social intelligence in an interactive system. Therefore, we drew up a dedicated questionnaire, based on our knowledge of existing questionnaires.

The online IPIP questionnaire (Goldberg, 1999a,b) offers a wealth of questionnaire items that assess personality and social traits in oneself. It is a questionnaire that primarily reflects the big five dimensions that are also addressed by the NEO-PI or NEO-FFT (Costa and McCrae, 1992). The difference is that the IPIP is in the public domain, and one is free to use any number of items to fit the purpose of the research being conducted. The IPIP items have the additional benefit that they have been compared and cross-validated with many existing questionnaires. A large number of items in the first version of the SBQ were taken from the IPIP pool and then rephrased to address the perception of social intelligence in an animated character.

The procedure for choosing certain items rather than others was as follows. Items were selected on the basis of scales that in some way reflected affective responses to others,

Table 1
Twenty scales in the SBQ

Altruism	Compliance	Gregariousness	Sociability
Amicability	Dutifulness	Helpfulness	Socialization
Assertiveness	Eagerness of effort	Likeability	Straight forwardness
Compassion	Empathy	Modesty	Trust
Competence	Good impression	Responsibility	Warmth

presenting oneself in a socially conventional way, and conveying an air of competence. Out of the search 172 items for 20 dimensions were selected. See Table 1 for the scales. Items that were in the relevant scales were chosen on the basis of whether they made sense in the context of assessing an animated character, and whether the items could be answered by respondents with someone else in mind, rather than oneself. The result was an average of about 8–9 items per scale. Although a questionnaire with a total of 172 items is considered average in the field of psychology, it is considered fairly large in the area of human–computer interaction.

Making a selection of items belonging to a scale, and turning it into a questionnaire of its own, does not ensure a good questionnaire; however, since the questionnaire was for the specific purpose of the study, there was no need for a large-scale validation. It was therefore decided to check for one sort of reliability: internal (scale) consistency. Cronbach's Alpha was calculated for each scale. A principal components analysis was also conducted to see if the distinctness of the scales was manifest in the resulting principal components.

5.1. Method

The questionnaire needed to serve the function of validating two distinct experimental states that we wanted to create in the study with the home dialogue system. In order to test whether the items we had gathered could function as a coherent questionnaire, we decided to put the questionnaire to the test in a study. This study would also show how to shorten and refine the questionnaire.

In light of the large number of items, the large number of participants needed, and the short time in which to conduct the study, it was decided to conduct the study online. For this reason it was not possible to let participants interact directly with a social or a non-social character. It was, however, possible to show participants situations in which an animated character displayed socially intelligent or socially unintelligent behaviour. Two 1-min movie clips were selected; one in which an animated character was clearly socially intelligent in some scenes (displaying empathy, gently trying to support a dying animal, sheltering a baby, displaying appropriate sorrow and restraint), and another in which the same character was devoid of any socially intelligent behaviour in other scenes (being unsusceptible to others' emotional state, unwittingly offending and picking a fight with two significantly larger and stronger animals, being unaware of the way another animal is trying to help him).

5.2. Design and procedure

A one-factor between subjects design was adopted, with two levels of social intelligence. Participants were invited by e-mail to go to the website with the movies where they saw either the movie clip in which the animated character acted in a socially intelligent way or the one where the character's conduct would be disastrous in terms of social intelligence. After viewing the video clip the participants were asked to fill in the questionnaire. They were instructed to answer the questions on the basis of the impression they had formed of the character. In practical terms this meant that the participants had to generalize, instead of just answer questions about the specific situation that was presented to them. The intention was to see if people could make a judgement about a character on the basis of a brief encounter.

All 172 items were formulated in the form of statements, like: 'The character…dislikes talking about himself', or: 'The character…trusts what others say'. The participants rated on a five-point scale (1, agree; 2, somewhat agree; 3, neither agree or disagree; 4, somewhat disagree; 5, disagree) the extent of their agreement or disagreement with the statement.

5.3. Results

Distinguishing experimental conditions. In total, 82 participants took part in the online study. The first analysis done was a Kruskal-Wallis one-way variance analysis to identify which scales the participants rate significantly differently in the two separate states.

Table 2 shows the means of the different scales for the two states and the *P* values. Out of 20 scales, 14 turned out to result in a significant difference between the two states. The overall mean of the scales also resulted in a significant difference. In short, there was a significant main effect of the state on several scales, and the overall mean of the scales. The final version of the SBQ only considered the items from the scales that distinguished between the two states.

Internal consistency. Internal consistency of all 20 scales was measured by calculating Cronbach's alpha (Table 2). In psychology, an internal consistency of 0.7 and higher is considered acceptable (Decoster and Claypool, 2004). A table of the alphas emerging from the analysis is given below.

On the basis of the alphas, we can say that three-quarters of the scales were consistent enough to be used in the questionnaire. In terms of shortening and refining of the questionnaire, what this means in concrete terms is that items from most scales could be used for the final questionnaire. There were really only two scales where the consistency was so poor that the corresponding items were not considered further.

Principal components analysis. A principle components analysis was performed as an extra check to see whether the scales that were used had, in fact, some common underlying components.

A five-factor solution emerged out of the analysis. When looking at content level at the items that loaded highly for the components it was possible to label them appropriately. Component one that explained 19.6% of the variance contained items that were indicative of someone 'who cares for others'. The second component with 10.6% explanation

Table 2
Means and significance level of the ANOVA

Scale	Socially intelligent *M*	Socially intelligent SD	Socially unintelligent *M*	Socially unintelligent SD	*P*	Cronbach Alpha
Altruism	2.171	0.787	2.756	0.726	0.001[a]	0.8212
Amicability	2.655	0.899	2.740	0.785	0.639	0.7754
Assertiveness	2.753	0.690	2.407	0.570	0.013[a]	0.7590
Compassion	2.130	0.714	2.366	0.598	0.180	0.7902
Competence	3.146	0.644	3.502	0.515	0.008[a]	0.7573
Compliance	2.829	0.605	3.251	0.554	0.002[a]	0.4930[b]
Dutifulness	2.705	0.740	3.442	0.681	0.000[a]	0.8345
Eagerness	3.063	0.714	3.015	0.734	0.690	0.8238
Empathy	2.497	0.525	2.957	0.581	0.001[a]	0.8494
Good impression	3.201	0.656	3.915	0.528	0.000[a]	0.7605
Gregariousness	1.967	0.773	1.863	0.543	0.813	0.8416
Helpfulness	2.217	0.640	2.939	0.698	0.000[a]	0.8158
Likability	2.085	0.605	2.433	0.577	0.011[a]	0.6839
Modesty	3.054	0.831	3.498	0.554	0.010[a]	0.8392
Responsibility	2.503	0.497	2.970	0.425	0.000[a]	0.5808[b]
Sociability	2.741	0.537	3.138	0.417	0.000[a]	0.6285[b]
Straightforwardness	3.011	0.437	3.237	0.320	0.005[a]	−0.4909[b]
Sympathy	2.236	0.793	2.902	0.867	0.001[a]	0.6615[b]
Trust	2.111	0.611	2.306	0.603	0.189	0.7959
Warmth	2.056	0.607	2.091	0.510	0.619	0.6853
General mean	2.544	0.407	2.867	0.329	0.000[a]	

[a] Significant with $\alpha=0.05$. Lower scores indicate higher perceived social intelligence.
[b] Cronbach alpha too small to fix.

variance can be termed ‘approaching others positively’. The third component ‘sure of own skills’ explains an additional 8%. The last two components ‘helpful’ and ‘cheerful’ explain 5.1 and 3%, respectively, of the variance, resulting in a total variance of 46.3% explained. Intuitively, these components can be thought of as the main indications of what is considered socially intelligent. The components also mostly contain items in the scales that distinguished to the maximum the two states in the Kruskal-Wallis variance analysis.

5.4. Selection of items for the SBQ

Taking all three analyses into account the following procedure was followed in order to reduce the number of items from 172 to 20 scales to make a ‘reasonably’ sized questionnaire that would distinguish between the two intended states to the maximum, as well as have scales that were consistent.

First of all, the items that loaded highly for the five principal components were thoroughly examined. From these items we tried to pick from the scales only those items that according to the Kruskal-Wallis analysis could distinguish between social and non-social states, and these items were also checked for internal consistency. A lot of items

Table 3
The final scales of the SBQ

Altruism	Modesty
Assertiveness	Responsibility
Competence	Sociability
Dutifulness	Sympathy
Empathy	Trust
Helpfulness	

could be removed in this way, and the best kept for the final version of the SBQ. The final version had 57 items, in 11 scales (Table 3). The items were in the form of statements like:

The robotic cat believes that others have good intentions.
The robotic cat is not interested in others' problems.
The robotic cat has a high opinion of itself.
The robotic cat tells the truth.
The robotic cat says inappropriate things.

Participants rated their level of agreement on a five-point scale.

It should be noted, however, that the means found were almost all in the socially intelligent direction. A five-point scale was used with '3' being 'neutral'. The two movie clips picked for the study were typical of a socially intelligent and a socially unintelligent character. However, the responses given by the participants are not in the socially unintelligent area. A few comments can be made about this result. It can be stated that the participants also found the socially unintelligent character somewhat social. However, we believe that the one-minute situation of the socially intelligent animated character was found to be funny rather than lacking in social intelligence. This 'funniness' could have also changed the judgements in favour of the character. Despite this problem that made the effect small, the fact remains that the questionnaire does distinguish between the two. The problems that made the effect small were not likely to occur in the experiment that was planned, for the obvious reason that the encounter would be longer, and short funny events would not occur during the entire session.

6. Experiment 2: home dialogue system

This study was situated in the HomeLab facility, which provides a laboratory simulation of a home setting. The HomeLab has an extensive observational infrastructure and is primarily used to test innovative technology in an almost naturalistic home environment. The aim of this study was to implement certain socially intelligent behaviours in the domestic robot iCat with the intention of finding out what the effect of these behaviours would be on the participants. Multiple methods were used to answer the questions we set.

6.1. Method

6.1.1. Participants

All 36 participants were recruited through an external agency and were paid for their participation. There were 15 women and 21 men. They all had at least some basic experience of using e-mail and the Internet. They were randomly assigned to one of the two experimental states.

6.1.2. Design

A one-factor between-subjects design was adopted with social intelligence being manipulated. Two states were created: A 'socially intelligent' state (condition 1) and a 'socially neutral' state (condition 2).

In the first state the robot spoke using synthesized speech from a text-to-speech engine. The speech was accompanied by lip-synchronization. The iCat blinked its eyes throughout the session, and displayed facial expressions and head movements. The behaviours-listed below-were guided by a list of aspects indicative of social intelligence.

- *Listening attentively*. By looking at the participant when s/he talks and occasionally nodding its head.
- *Being able to use non-verbal cues displayed by the user*. Responding verbally to repeated wrong actions on the part of the participant by offering help.
- *Assessing well the relevance of information to a problem at hand*. By stating what is going wrong before offering the correct procedure.
- *Being nice and pleasant to interact with*. By staying polite, mimicking facial expressions (smiling when participant smiles, for example), being helpful.
- *Not ignoring affective signals from the user*. By responding verbally or by displaying appropriate facial expression to obvious frustration, confusion, or contentment.
- *Displaying interest in the immediate environment*. The immediate environment being the participant and the equipment used in tasks, by carefully monitoring the person and the progress of the tasks.
- *Knowing the rules of etiquette*. By not interrupting the participant when s/he is talking.
- *Remembering little personal details about people*. Addressing the participant by name, remembering login information, and passwords if asked.
- *Admitting mistakes*. By apologizing when something has gone wrong, but also when no help can be provided upon participant's request.
- *Being expressive*. By showing facial expressions while talking, if appropriate.
- *Thinking before speaking and doing*. By showing signs of thinking (with facial expression) before answering questions or fulfilling the participant's request.

The facial expressions and head movements were pre-programmed. The expressions were blocks of certain behaviours with different intensities. For example, there were four pre-programmed expressions for 'smiling', each at a different intensity. The experimenter would type in responses to the participant, although there were a few pre-set responses for situations that often occurred. For example, the instructions for working the DVD recorder were pre-fabricated. The experimenter would initiate these pre-programmed social behaviours at appropriate moments during the sessions (Wizard of Oz).

Condition 2 was the 'socially neutral' state. The iCat did not display any facial expressions and did not blink its eyes. It talked and used lip-synchronization. It responded verbally only to explicit questions from the participant. The aspects of social intelligence listed above, were not supported. The only help provided without the explicit request by the participant would be in cases where they got stuck to such an extent that they would not be able to continue the experiment without help.

6.1.3. Tasks

Participants were asked to perform two tasks. The first task was to program a DVD recorder to record three broadcast shows for the upcoming week. This task was intended to let participants become familiar with having a robotic cat standing on the table that they could talk to in order to get support when operating the DVD recorder.

The second task was an online auction. A special auction site was made for the purpose of this study. The task was to register and buy several items on a list. For registration as a new user, the site required a valid web-accessible e-mail account. Participants could also give their e-mail details (login and password) to the iCat if they wanted it to monitor their items. The rationale behind the choice of the auction task was that we wanted participants to be immersed in an intensive task, while in the background the iCat was always available for help. The task was also designed to convey the ability of the home dialogue system to participate in the task by accessing sensitive information if authorized. Instead of artificial costs we selected a task that required people to give the actual password of their e-mail system in order to let the iCat provide support during the auction task. This was done to make the task more personal and to increase the participant's commitment to it.

6.1.4. Measures

Social behaviors questionnaire (SBQ). The newly developed SBQ was used to test whether participants rated the two experimental states as different.

User satisfaction questionnaire (USQ). This is an in-house instrument that helps assess user satisfaction with consumer products (De Ruyter and Hollemans, 1997). The USQ was used to assess the level of satisfaction with a DVD recorder that participants had to operate during the experiment. Due to the effects of social intelligence in the larger cognitive context (including attitudes towards the DVD recorder), it is plausible that the quality of the interaction with the iCat may influence how participants experience the interaction with the DVD recorder. It was expected that satisfaction with the DVD recorder would be higher in the socially intelligent state.

The unified theory of acceptance and the use of technology (UTAUT). The UTAUT (Venkatesh et al., 2003) is a measure of technology acceptance in the workplace. It is used to evaluate the likelihood that new technology introduced in industry will be accepted by the employees who will be required to use it. In our study the questionnaire was used to measure the extent to which participants would use iCat at home after the experiment. Given the difference of the application domain, a revised version of the UTAUT was used. Again, because social intelligence was expected to affect more than just direct emotions, it was hypothesized that technology acceptance would be higher in the state where the iCat acted as socially intelligent.

6.1.5. Rating of own performance

In a post-experimental questionnaire participants could indicate on a five-point scale what they thought about their own performance during the experiment.

6.1.6. Observations

Finally, we noted the number of times that participants asked the robot general questions and the number of times they asked questions about the experimental tasks. We also noted the number of times that participants looked at the robot during the entire session.

This multiple set of measures was designed to test both the direct effects of iCat's behaviours and the potential implicit spillover effects, like satisfaction with the DVD recorder.

6.2. Procedure

Participants were welcomed and it was explained that they were going to do two tasks. They were also told that while they did those tasks there would be a robot cat on the table that could be addressed if they needed or wanted its help. There would be times that the robot would initiate conversation when it thought it might be able to help This is all the information the participants were given about operating the robot. They were not informed about any functionalities that iCat had. The manner of instruction was such that emphasis was placed on completing the tasks, whereas interaction with iCat was secondary.

Next, the participants were given a task-booklet with instructions on the two tasks and were brought into the living room. During the experiment they were left alone in the living room (Fig. 2 shows the test setting with the participant and the iCat). They were given 10 min to program the DVD recorder to record three shows of their choice. The second task was the online auction task. There was a laptop in the living room that had a broadband connection to the Internet. Participants could access the auction by double-clicking a shortcut on the desktop. Before bidding on items they were required to register as a new user with a valid web-accessible e-mail account. After that they were required to bid on and acquire several items.

The participants were also told that simply bidding on an item would not ensure acquiring the item. 'Others' on the web were also bidding and they could be out-bid by these others. Notifications of higher bids were sent to the e-mail account that participants used to register. The e-mail account had to be monitored if the participants wanted to complete the task successfully. When the opportunity arose, the iCat could monitor their e-mail account for 'outbids' if the participants authorized it to do so. The iCat was there to help in many other ways as well. If participants could not manage to register as a new user, the iCat could register on their behalf. This was done in both states; in the socially neutral state this was done only when participants did not succeed in registering within 12 min. The iCat could also give information on the items that were offered on the auction.

Fig. 2. Snapshot from the test sessions showing: the participant (LIV1), an overview of the living room (LIV2 and LIV3) and a close up of the iCat.

If authorized, it could place bids for the participants. The participants were allowed 20 min to buy the listed items.

After this the participants were taken to a separate room. They gave their first impressions of the experiment in an interview. Then, they filled in the three questionnaires in the sequence SBQ, USQ and UTAUT. Finally, the participants were interviewed on their performance in the auction.

6.3. Results

6.3.1. Questionnaires

In light of the limited number of participants and the questionnaires having ordinal rating scales, Kruskal-Wallis one-way variance analyses were performed to see whether the responses to the three questionnaires differed from each other in the two states. An $\alpha = 0.05$ was set for all analyses.

SBQ. For the SBQ the difference between the two states was significant ($\chi^2 = 5.938$, df = 1, $P < 0.05$). Inspecting the means indicated that our hypothesis was confirmed: participants thought the socially intelligent iCat was indeed more socially intelligent than the neutral iCat.

USQ. The difference in evaluating the DVD recorder was also significant ($\chi^2 = 4.294$, df = 1, $P < 0.05$). Participants who interacted with the socially intelligent iCat, were more satisfied with the DVD recorder.

UTAUT. The responses to the UTAUT (Table 4 shows the means of the questionnaires) also resulted in a significant difference ($\chi^2 = 9.633$, df = 1, $P < 0.05$). Participants who worked with the neutral iCat were less inclined to want to continue working with iCat at home.

Table 4
Means of the questionnaires

Questionnaire	Socially intelligent	Neutral
SBQ	1.98	2.34
USQ	1.92	2.49
UTAUT	1.71	2.33

6.3.2. Interview

From the semi-structured interview it is clear that there was no difference between how participants evaluated their performance in the auction. There were rating-scales attached to the questions asked. On a scale from 1 (not pleased with performance) to 5 (very pleased with performance), the average for both states was 3.8 ($\chi^2 = 0.170$, df = 1, $\alpha = 0.680$).

At the end of the session people also had the opportunity to talk freely about the experiment. This also enabled each participant to answer the following question: 'If you had iCat at home, what would you like it to do for you there?' There were responses like 'all the electronics at home'. Examples mentioned included the more obvious controlling the lights, heating system and entertainment equipment like TV, VCR and DVD player, but also less obvious things like the microwave, oven, and toaster were named. Some, however, also mentioned things like using iCat as a filter for Internet and TV programmes for their children and having iCat as a cooking (recipe) guide in the kitchen. More private tasks were also noted, like having their e-mail checked for them, screening telephone calls, and Internet banking. No differences were found in the pattern of responses of the participants in the different states. There was, however, a difference in the constraints imposed before authorizing iCat to access this personal information. Of the 11 participants' in the socially intelligent state that would like iCat to handle their private tasks, six would like some market research data that tells them that it is safe and secure to use iCat for these things. The other five would use iCat without further evidence. Participants in the neutral condition felt differently. There were 12 people who would authorize iCat for personal tasks. Four of them would like research evidence before using iCat. Only one person said he would let iCat deal with personal applications immediately. But seven people were not sure if research would be enough. They wanted to experience iCat further before allowing it extended access to their private information. They stated that they would first give it small tasks. Only over the longer term, after proven success, would they give it full authorization. Only one of the participants would use iCat as it is now.

6.3.3. Observations

We counted the times that participants asked iCat questions. The Wilcoxon–Mann-Whitney test was used to analyze the data. The averages in both states were close: 13.6 times in the socially intelligent versus 11.1 in the neutral state ($Z = -0.954$, $P > 0.05$). Additionally, up to 4.9 questions on average were posed about items in the auction in the socially intelligent state. In the neutral state this average was 3.2 ($Z = -0.486$, $P > 0.05$). Although at first glance there seemed to be a difference between the numbers of times that participants looked at iCat (11.6 and 6.0 for socially intelligent and neutral condition, respectively), it was not significant ($Z = -1.134$, $P > 0.05$). In many cases participants looked at the robot in anticipation of an answer.

7. Discussion

The results from the SBQ verify the distinctness of the experimental states that we wanted to create: participants rated the socially intelligent iCat as more social than the neutral one, which seems to validate that the collection of behaviours implemented or simulated in the iCat do help it to exhibit social intelligence.

The USQ also had a differential effect between the two states. Since the USQ was developed to test satisfaction with a consumer product after thorough interaction with that product and the DVD recorder task only consisted of exploring one function in a time frame of 10 min, the significant difference found between the two experimental states is quite remarkable. Apparently the positive experience of the interaction rubbed off on the perception of working with the DVD recorder. It should be noted that for filling in the USQ, the participants were asked to ignore the help from iCat as much as possible: they were requested to strictly evaluate the DVD recorder.

Given the fact that we used a modified version of the UTAUT, we can only draw tentative conclusions from this measurement. Five out of the six scales were significantly different between the two states. The two scales in the original UTAUT that were removed could not have changed the outcome. The UTAUT was applied to the iCat and, as such, it shows the explicit positive effect of social intelligence manipulation.

There was no significant effect regarding perceived auction performance; most participants thought they did pretty well in both states. Being regular Internet users they were familiar with auction sites like ebay. The task therefore did not pose problems and they felt they did very well. With hindsight however, they indicated that they would have liked to ask the iCat more questions regarding the products in order to decide faster which products to bid on. The participants would also have liked to delegate more chores to iCat. Not all the participants, for example, discovered that they could ask iCat to keep an eye on their bids and have them notified when they were out-bid on an item. The 83% who did discover this function did delegate. Some participants tested iCat to the extent that they asked it to place counter offers when someone outbid them. They would also have liked iCat to give more reasons for recommending products. The participants that were not very satisfied with how well they had performed were those who in their daily lives do not spend much time on the web or on the computer (5 participants). They had basic experience of using e-mail and the Internet, but were not as proficient as most of the other participants. Most likely, dissatisfaction was caused by the difficulty of navigating in unfamiliar territory, namely, an auction site. These same participants also had trouble registering as a new user, because they did not have their e-mail passwords readily available or gave e-mail accounts that were not web-accessible (they were accustomed to using a once installed e-mail client on their PC at home and did not know how to access their accounts via the web).

The overall impression was that participants were more ‘social’ with the socially intelligent iCat: they were much more inclined to laugh and divert conversation to a relational level. One participant, for example, was making conversation about the plasma TV set in the living room, commenting on how great it would be to watch soccer on such a cool screen, and whether iCat thought this living room would be available for rental that evening. Another person asked if iCat minded if he left the TV on to listen to some music while doing the task and continued by asking what type of music iCat enjoyed listing to. Participants also asked for more details on questions, than they did with the neutral iCat. They were more curious about the reasons why the social robot said the things it said than when they were interacting with the neutral robot. For example, when asked which LDC monitor was a good one (to buy in the auction task) they were happy that iCat could help by naming a product. But they were curious how it knew this and why it was the best. They were also more inclined to ask iCat's opinion on the other LCD monitors. They asked these questions politely and using full sentences. In the case of the neutral iCat, they were more inclined to take the suggestion of the best LCD monitor for what it was and not continue to probe further.

Participants in the socially intelligent state liked the fact that the robot was expressive in terms of facial expressions. They liked the fact that it nodded and shook its head in response to their talking. Overall, they agreed that it was a robotic cat with a face and it was only natural that it used its full potential this way. It made the robot friendlier and easier to approach. On the other hand, participants in the neutral state, who experienced only the talking and the lip-synchronization while talking, also liked iCat in that way. After all, they argued, it is a robot and it should not try or pretend to be anything other than that. Moving and facial expressions, according to them, would only look like a poor attempt to seem alive and it would likely annoy and distract you from whatever you are doing. This finding shows how hard it can be to imagine and evaluate something that has not been experienced.

Although many participants were reticent about disclosing their e-mail details to the robot, many said that they would not have this problem at home. It would be a wonderful thought to not have to turn on the computer just to check e-mail, but that a robot could do that for you. At the very least, participants would like a robot to inform them whether or not they have new e-mail messages in their mailbox. For many of them, it is also important to know who the e-mails are from and the subject. Some even wanted to leave the reading of the e-mail to the robot as well. Participants were also prepared to delegate other activities, like online banking to the robot. Although they did expect to be kept informed about its activities. We note here that because of the experimental set-up, participants seemed to attribute quite extensive abilities to the robot. Clearly, their PC is equally capable of 'reading' their e-mail, or notifying them of new messages; however, during the discussions it turned out that the apparent perceptive abilities of the robot seemed to equate it to an ever perceptive and automatic computational environment. In this light, their inclination to trust the robot with private information seems quite remarkable. Although we did not explicitly focus on trust as a concept, the study does suggest a relationship with social intelligence. In fact, the whole positive attitude towards the socially intelligent iCat points to a 'halo effect'. This might include more affective concepts linked to social intelligence.

8. Conclusion

The initial research question that we set was whether a layer of social intelligence in a home dialogue system could elicit positive attitudes towards the technology embedded in the home environment. The results reported suggest that the implementation of socially intelligent behaviour in a home dialogue system could have this effect. Participants in this study had the experience of working with a robot that could 'see' what the participant was doing, whether this was a task on the TV and DVD recorder or on the Internet. Although the participants were well aware of the level of invasion technology like this would have on their lives—should it be allowed into their homes—many of them welcomed this fact. The prime reason seems to be that users do not enjoy the chores brought about by the proliferation of technology in their surroundings and they would happily delegate these chores to a robot in the home.

Our study showed that a few social behaviours in a robot are enough to remove a lot of the discomfort that is brought about when moving interactive systems into the background. One plausible explanation is that participants are more inclined to accept a single centralized interface for an Ambient Intelligence environment rather than a distributed set of separate products. Implementing some thoughtful aspects of social intelligence in a perceptive robot, so that it is able to understand the problems that users experience, makes the robot easier to communicate with and more trusted by users.

While most research on social robotic characters has concentrated on the interaction with the robot as the focus of attention, this study focused on the role of a robot as a home dialogue system. The robotic interface here served as a tool to affect changes in the environment and accomplish tasks. The interaction with the iCat was not the participants' priority. Despite its background function, the iCat and the behaviours that it displayed had a significant effect on the level of satisfaction with the embedded systems, acceptance of the technology, and sociability towards the system. This study opens an avenue for future research on the interaction between humans and Ambient Intelligence technology. The concept of social intelligence is important not just for direct interaction with robotic or even screen characters, but has relevance in systems that do not necessarily have a social function.

Regarding related research into social intelligence of on-screen and robotic characters, this study has made a threefold contribution.

- We have demonstrated through our design how a collection of human-like behaviours can lead the character to be perceived as socially intelligent.
- With the social behaviours questionnaire we developed, we have provided a means to evaluate this perception.
- We have demonstrated the relevance of social intelligence as a concept for studying the interaction between humans and computational characters. Researchers often assume this relevance implicitly; by showing how an increase in perceived social intelligence impacts positively on people's perceptions of a system, we can substantiate this assumption.

On the basis of the results we have presented, future research will be able to explore the most effective ways of achieving social intelligence and forge links between lower-level behaviours of robots or on-screen characters and the resulting perception of social intelligence.

References

Aarts, E.H.L., Harwig, R., Schuurmans, M., 2001. Ambient intelligence. In: Denning, P. (Ed.), The Invisible Future. McGraw Hill, New York, pp. 235–250.

Banham, K.M., 1968. Social Competence Inventory for Adults: A Social Competence Inventory for Older Persons. Family Life Publications, Durham.

Bartneck, C., 2002. EMuu – an embodied emotional character for the ambient intelligent home. Ph.D. Thesis Eindhoven University of Technology.

Bartneck, C., 2003. Interacting with an embodied emotional character, Proceedings of the DPPI2003 Conference, Pittsburgh.

Beskow, J., McGlashan, S., 1997. Olga – a conversational agent with gestures, In Proceedings of the IJCA'97, 1997, Workshop on Animated Interface Agents – Making Them Intelligent.

Bickmore, T., Cassell, J., 2001. Relational agents: a model and implementation of building user trust, Proceedings. ACM CHI 2001, Seattle, Washington, pp. 396–403.

Bickmore, T., Picard, R.W., 2004. Towards caring machines. CHI Extended Abstracts 2004 2004;, 1489–1492.

Breazeal (Ferrell), C., Scassellati, B., 2000. Infant-like social interactions between a robot and a human caretaker. In: Dautenhahn, K. (Ed.), Special Issue of Adaptive Behavior on Simulation Models of Social Agents.

Breazeal, C., Hoffman, G., Lockerd, A., 2004. Teaching and Working with Robots as a Collaboration. In Proceedings of the Third International Joint Conference on Autonomous Agents and Multi Agent Systems (AAMAS'04), 1030–1037, New York, NY.

Bruce, A., Nourbakhsh, I.R., Simmons, R.G., 2002. The role of expressiveness and attention in human–robot interaction. ICRA 2002 2002;, 4138–4142.

Costa Jr., P.T., McCrae, R.R., 1992. Personality NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) Professional Manual. Psychological Assessment Resources, Odessa, FL.

Decoster, J., Claypool, H.M., 2004. A meta-analysis of priming effects on impression formation supporting a general model of informational biases. Personality and Social Psychology Review 8 (1), 2–27.

De Ruyter, B.E.R., Hollemans, G., 1997. Towards a User Satisfaction Questionnaire for Consumer Electronics: Theoretical Basis. Natuurkundig Laboratorium Philips Electronics N.V., Eindhoven. NL-TN, 406/97.

Ford, M.E., Tisak, M.S., 1983. A further search for social intelligence. Journal of Educational Psychology 75, 196–206.

Goldberg, L.R., 1999a. A broad-bandwidth, public-domain, personality inventory measuring the lower-level facets of several five-factor models. In: Mervielde, I., Deary, I., De Fruyt, F., Ostendorf, F. (Eds.), Personality Psychology in Europe, vol. 7. Tilburg University Press, Tilburg, The Netherlands, pp. 7–28.

Goldberg, L.R., 1999b. http://ipip.ori.org/ accessed February 15, 2004.

Gough, H.G., 1968. Chapin Social Insight Test Manual. Consulting Psychologists Press, Palo Alto, CA.

Heylen, D., Van Es, I., Van Dijk, E.M.A.G., Nijholt, A., 2004. Experimenting with the gaze of a conversational agent. In: van Kuppevelt, J., Dybkjaer, L., Bernsen, N.O. (Eds.), Natural, Intelligent and Effective Interaction in Multimodal Dialogue Systems. Kluwer Academic Publishers, Dordrecht.

Hüttenrauch, H., Eklundh, K.S., 2002. Observations from a long-term user study with a service robot, In Proceedings of the 11th IEEE International Workshop on Robot and Human Interactive Communication, Roman'2002, pp. 158–163.

Imai, M., Ono, T., Ishiguro, H., 2001. Physical relation and expression: joint attention for human–robot interaction. Robot and Human Interactive Communication 2001. Proceeding 2001;, 512–517.

Keating, D.K., 1978. A search for social intelligence. Journal of Educational Psychology 70, 218–233.

Markopoulos, P., 2004. Designing ubiquitous computer human interaction: the case of the connected family. In: Isomaki, H., Pirhonen, A., Roast, C., Saariluoma, P. (Eds.), Future Interaction Design. Springer.

Moos, F.A., Hunt, K.T., Omwake, K.T., Ronning, M.M., 1927. Social Intelligence Test. The Center for Psychological Service, Washington, DC.

NEC: http://www.incx.nec.co.jp/robot/PaPeRo/english/p_index.html, accessed June 10, 2004.

Norman, D.A., 1998. The Invisible Computer. MIT Press, Cambridge, MA.

Omron Corporation. http://www.necoro.com/newsrelease/ accessed August 1, 2004.

Pentland, A., 2000. Perceptual intelligence. Communications of the ACM 43 (3), 35–44.

Reeves, B., Nass, C., 1996. The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places. Cambridge University Press, Cambridge.

Sidner, C.L., Kidd, C.D., Lee, C.H., Lesh, N.B., 2004. Where to look: a study of human–robot engagement, ACM International Conference on Intelligent User Interfaces (IUI) 2004 pp. 78–84, January 2004.

Sternberg, R.J., Smith, C., 1985. Social intelligence and decoding skills in nonverbal communication. Social Cognition 3, 168–192.

Strang, R., Brown, M.A., Stratton, D.C., 1942. Test of Knowledge of Social Usage. Columbia University, New York.

Thórisson, K.R., 1997. Gandalf: an embodied humanoid capable of real-time multimodal dialogue with people. First ACM International Conference on Autonomous Agents, Marina del Rey, California, February 5–8, pp. 536–537.

Van Breemen, A.J.N., 2004. Bringing robots to life: applying principles of animation to robots. CHI2004 workshop Shaping Human–Robot Interaction, Vienna.

Venkatesh, V., Morris, M.G., Davis, G.B., Davis, F.D., 2003. User acceptance of information technology: towards a unified view. MIS Quaterly 27 (3), 425–478.

Vernon, P.E., 1933. Some characteristics of the good judge of personality. Journal of Social Psychology 4, 42–57.