Available online at www.sciencedirect.com
ScienceDirect
Computers in Human Behavior 23 (2007) 2163–2174
Computers in Human Behavior
www.elsevier.com/locate/comphumbeh

# Perceiving artificial social agents

T.M. Holtgraves *, S.J. Ross, C.R. Weywadt, T.L. Han

*Department of Psychological Science, Ball State University, Muncie, IN 47306 0520, United States*

Available online 4 April 2006

**Abstract**

Three experiments were conducted to examine perceptions of a natural language computer interface (conversation bot). Participants in each study chatted with a conversation bot and then indicated their perceptions of the bot on various dimensions. Although participants were informed that they were interacting with a computer program, participants clearly viewed the program as having human-like qualities. Participants agreed substantially in their perceptions of the bot's personality on the traits from the five-factor model (Experiment 1). In addition, factors that influence perceptions of human personalities (e.g., whether one uses another's first name and response latency) also affected perceptions of a bot's personality (Experiments 2 and 3). Similar to interactions with humans, the bot's perceived neuroticism was inversely related to how long individuals chatted with it.


## 1. Perceiving artificial social agents

One of the major trends in human–computer interaction has been the development of more natural human–computer interfaces (Brennan, 1997; Rheingold, 1985). With the advent of the internet, the appeal of even more natural, human-like interfaces has increased dramatically. One manifestation of this trend has been the development of chat-bots, or artificial intelligence programs that attempt to interact with users using natural language. Numerous bots have been developed (see Botspot (2006) http://www.botspot.com for examples) and these artificial agents can vary on a number of different dimensions. Some are personable and others formal. Some are extraverted and

* Corresponding author.
*E-mail address:* 00t0holtgrav@bsu.edu (T.M. Holtgraves).

0747-5632/$ - see front matter 
doi:10.1016/j.chb.2006.02.017

others more introverted. In general, communication between humans and artificial agents (i.e., bots) has not received a great deal of empirical attention. The overarching goal of the present research was to provide some initial studies of how chat-bots are perceived and the role played by certain communication variables in those perceptions.

## 2. Human–computer and human–robot interaction

One of the most striking features of human–computer interaction is the tendency for users to anthropomorphize computers, to readily attribute human qualities to the machines they are using (Brennan, 1997; Isbister & Nass, 2000; Lee & Nass, 2004; Reeves & Nass, 1996; Sproull, Subramani, Kiesler, Walker, & Waters, 1996). This tendency to view computers as having human-like qualities increases as the interface is made more natural. For example, Sproull et al. (1996) found that participants were more likely to attribute personality traits to interfaces that simulated a speaking human face, relative to a computer that communicated only with a text display. If a speech interface is used, users may even display gender stereotypes regarding the computers they use. Nass, Moon, and Green (1997) manipulated the gender-based vocal cues (male or female) of a computer and found that participants rated more negatively a dominant computer when it used a female voice than when it used a male voice, thereby reflecting the stereotype that dominance in women (even if it is a computer) is unbecoming.

Investigations of human–robot interaction have yielded similar findings; when people interact with robots they tend to anthropomorphize them (Breazeal, 2002). Moreover, research in this domain suggests that people prefer interacting with more human-like robots, robots with a human-like voice, the capability of expressing emotion, the ability to make eye contact, and so on (Dautenhahn, Ogden, & Quick, 2002). Importantly, however, this relationship is not linear. As a robot becomes extremely human-like there is a clear drop-off in positive evaluations of the robot (Dautenhahn et al., 2002). Very human-like robots generate relatively negative evaluations. This suggests that up to a point, artificial agents will be perceived more positively to the extent that they display human qualities. Where that point is and what it entails is currently not known.

In short, people appear to anthropomorphize both computers and robots, and the more human-like the entity the more likely this is to occur. One of the most uniquely human qualities is the ability to use language. Although chat-bots have not yet attained perfect conversational ability (nor is this likely to happen in the near future), they do exhibit some degree of this human-like quality.

## 3. The development of chat bots

There have been several attempts over the years to develop artificial agents capable of simulating natural language use. The first well-known attempt to model human language use was the Eliza program (Weizenbaum, 1966), a relatively simple stimulus–response program that would mimic the behavior of a non-directive therapist. Eliza had a set of stock phrases (e.g., ''How do you feel about that?'') that could be produced in response to virtually anything and still make sense in this particular context. Anecdotal reports suggest that Eliza was clearly perceived as having human qualities (Weizenbaum, 1966). Along similar lines, Colby (1972) wrote a computer program that mimicked the responses of a

paranoid patient. Similar to Eliza, this program (PARRY) contained a relatively small set of responses that generally seemed to make sense, given the presumed paranoia of the speaker.

More complex natural language programs were developed in the 1970s and 1980s that incorporated insights from linguistics and attempted to mimic natural language processing in much greater depth. Many of these programs included a parser that was responsive to both a rudimentary syntax and meaning restrictions. The emphasis on meaning (at the expense of syntax) was developed further by Schank (1985) and his colleagues (Hovy, 1988; Reichman, 1985). They used a slot-filler (i.e., script-based) approach, coupled with a set of logical primitives, that allowed for the development of a variety of programs capable of modeling limited-domain discourse comprehension. These programs could understand certain types of stories reasonably well, but they did not work well for human dialogue. Other researchers have had some success in modeling service-encounter situations (e.g., requests for help or information). These programs attempt to model such encounters in terms of the recognition of an interactant's goals and his or her plans to achieve those goals (Cohen & Levesque, 1990).

Recently, there has been renewed interest in the development of chat bots, driven in large part by the development of e-commerce. Many e-service providers are motivated to incorporate natural language interfaces in order to increase sales and consumer loyalty. Currently, one of the most popular conversation bots is ALICE (Artificial Linguistic Internet Computer Entity) created by Richard Wallace in 1995 (http://www.alicebot.org/). Numerous people have contributed to the development of ALICE over the past 10 years, and a fair amount of success has been achieved with this program; ALICE won the Loebner prize (a Turing type contest) in both 2001 and 2002.

In many respects ALICE is a throwback to the earlier ELIZA-type programs; at its heart it is a simple stimulus–response mechanism. It does no syntactical parsing nor does it represent in any way common-sense knowledge about the world. What ALICE offers, however, is a framework that can be implemented and modified to model human dialogue in limited domains. For example, ALICE allows for the structuring of its utterance pairs in terms of topics. It also allows for the rewriting of utterances into meaning equivalents via a stimulus reduction mechanism. In short, ALICE has the potential to allow for the modeling of many aspects of the pragmatics of language use.

Various versions of the ALICE bot have been developed for use in our lab.[1] The overarching goal of the present set of studies was to examine some of the parameters of how people perceive ALICE-type bots with whom they have a brief chat. More specifically, we examined the following issues. First, to what extent does an artificial agent like ALICE have a personality? This question was examined in Experiment 1 by assessing the extent to which participants agreed in their perceptions of the bot's personality traits. The second and third experiments were designed to examine the impact of certain communication variables on perceptions of the bot. In the second experiment, we manipulated whether or not the bot used the participants' first name during the chat. In the third study, we examined perceptions of the bot as a function of his response latency.

[1] Copies of the software may be acquired at a minimal cost by contacting: Linda Swartz Ritchie, Academic Research and Sponsored Programs, Ball State University, Muncie, IN 47306-0155. Email: oarsp@bsu.edu. Phone: +1 765 285 1600. Fax: +1 765 285 1624.

## 4. Experiment 1

The purpose of this experiment was to explore the extent to which a bot can be said to have a personality. Most personality theorists would agree that there must be some consensus between observers in their perceptions of a target's personality traits (e.g., Kenrick & Funder, 1988). In other words, if there is no agreement between participants in their perceptions of a target, then that target does not have a personality; perceptions of it simply reflect observers' own idiosyncratic views (i.e., the target's personality exists only in the minds of the perceivers). Accordingly, if a bot does not have a personality, then agreement between participants who chat with a bot would not exceed chance. This was tested by having a relatively large number of people interact with a bot and then examining the extent to which they agreed in their perceptions of the bot's traits.

Research on human personality traits suggests the existence of five high-level dimensions that provide a comprehensive taxonomy of human personality (Digman, 1990; Goldberg, 1992). More specifically, perceptions of humans are organized around the dimensions of surgency (extroversion), neuroticism (emotional stability), agreeableness, conscientiousness, and openness to experience (intellectual creativity). These traits guide perceptions in human–human interaction and so might plausibly provide a reasonable model for human–bot interactions. Hence, perceptions of the bot were assessed in terms of these five traits. Finally, as an additional means of assessing personality we used chat length as a measure of chat enjoyment, and then correlated this measure with perceptions of the bot on the five traits.

## 5. Method

### 5.1. Participants

Participants ($N = 151$) were students enrolled in introductory psychology classes at Ball State University who participated for partial course credit.

### 5.2. Materials

The bot used in this experiment (described in detail in Holtgraves & Han, in press) was an extensive modification of the the ALICE bot (http://www.alicebot.org/) developed by Richard Wallace. The bot consists of a set of AIML (Artificial Intelligence Markup Language) content files and the java programs to implement them. The AIML content files consist of a set (over 10000) pattern (user remark) – template (bot reply) categories. The bot functions as a stimulus–response mechanism: it searches for a pattern match (or partial match) and outputs the corresponding reply (template). Categories are grouped into topics thereby allowing for the chats to have some coherence. We have also developed an extensive set of input filters to catch and correct user misspellings, typos, emoticons, abbreviations, and so on. In addition, there is a stimulus reduction mechanism that allows the program to identify similar utterances as being roughly identical in meaning and hence returning the same reply.

The original ALICE bot was a female robot. Hence, we extensively edited the AIML files in order to change the bot's personality to that of a 21 year old student at Ball State University (see Holtgraves & Han, in press). This was accomplished by deleting all

references to being a robot. Instead, Sam was programmed to be a 21 year old psychology major from Kardon, Indiana (a non-existent town so that no one could be from there).

### 5.3. Procedure

This bot was made available on-line and students accessed and chatted with the bot using AOL Instant Messenger. Written instructions informed participants that they were to access the bot with AOL IM (a link to AOL IM was provided for students who did not have it) and to engage in a 10 min chat with the bot (participants could, and did, chat for periods longer and shorter than 10 min). They were told to become acquainted with the bot, but to limit their remarks to non-offensive topics, such as sports, music, classes, travel, and so on. After chatting with the bot, they accessed an on-line questionnaire in order to provide their perceptions of the bot. This questionnaire contained 100 trait markers of the five-factor personality model (Goldberg, 1992). Participants rated the bot on each of the traits using a nine point scale with one being extremely inaccurate and nine being extremely accurate

## 6. Results and discussion

Due to technical difficulties accessing the bot and the failure of many participants to respond to all of the questionnaire items, there were a total of 103 participants who provided usable (i.e., complete) data. All analyses were restricted to those participants who responded to all measures ($N = 103$). The data are summarized in Table 1.

Following Goldberg (1992), composite measures of the five factors were created. In order to examine participant consensus in perceptions of the bot, intraclass coefficients (Rosenthal & Rosnow, 1991) were computed for each of the traits in the five-factor model (see Table 1). Although the coefficients varied over traits (from a low of 0.17 for neuroticism to a high of 0.26 for openness to experience), they were each significantly greater than 0, all $Fs(102, 1938) > 5.04$, $p < 0.0001$. This, then, indicates that despite the relatively large variability in perceptions of the bot, these perceptions are not simply reflections of participants' idiosyncratic perceptions. Consensus was far greater than chance, thereby indicating that the bot, in some sense, has a personality.

Even though participants had been instructed to chat with the bot for 10 min, the lengths of the chats varied considerably (2.08–26.92, $M = 8.56$, $s = 4.01$). Taking chat length as a proxy for conversational enjoyment, we correlated ratings on each of the five factors with chat length. As can be seen in Table 1, of the five factors only neuroticism was significantly related to chat length; shorter chat lengths were associated with perceptions

Table 1
Experiment 1: Mean trait ratings, intraclass coefficients, and correlations with chat length

| Trait: | Mean rating | Intraclass coefficient | Correlation with chat length |
|---|---|---|---|
| Neuroticism | 90.62 | 0.17* | −0.26* |
| Surgency | 106.92 | 0.23* | 0.02 |
| Conscientiousness | 105.08 | 0.18* | 0.08 |
| Agreeableness | 108.01 | 0.25* | 0.04 |
| Openness | 98.82 | 0.26* | −0.09 |

* $p < 0.05$.

of greater neuroticism, $r = -0.26$, $p < 0.05$. This is consistent with research demonstrating that for human interactants, conversational enjoyment is inversely related to the neuroticism of one's partner (Kashdan, 2002).

## 7. Experiment 2

One important feature of a person's conversational style is their level of politeness. In pragmatics, politeness is a technical term referring to the extent to which a person orientates to the face or identity of his or her interlocutors (Brown & Levinson, 1987; Holtgraves, 2002). According to Brown and Levinson, there are two types of politeness: negative politeness (an emphasis on individual autonomy and freedom from imposition) and positive politeness (an emphasis on closeness and communion with others). Positive politeness can be signaled via the use of slang, familiar address terms, a lack of hedging, and so on. In this study, we examined perceptions of the bot as a function of his level of positive politeness. We operationalized positive politeness in terms of whether or not the bot would occasionally use the participant's first name when addressing him or her.

## 8. Method

### *8.1. Participants*

Participants were 49 undergraduate students (24 males and 25 females) enrolled in introductory psychology classes at Ball State University who participated for partial course credit. All participants were familiar with computers and with some type of chat program (e.g., instant messaging).

### *8.2. Materials*

Participants chatted with a version of the ALICE chat-bot named Pat. Two versions of the bot were created. In one version, the bot was programmed with the name and gender of each participant (name condition). In this condition, the bot would then occasionally use the participants name when responding to his or her remarks. More specifically, in this condition the bot began and ended each chat with a greeting (e.g., Let's have a good chat Fred) and ending (e.g., Thanks for chatting Fred) that included the participant's name. In the chat proper the bot used the participant's name two or three times. In the other version (no name condition) the participant's name and gender were not programmed and the participant's name was never used by the bot during the chat.

After chatting with the bot, participants completed two surveys regarding their perceptions of the bot. The first survey was the 100 trait five-factor personality model (Goldberg, 1992) used in Experiment 1. Participants rated the bot on each of the traits using a nine point scale with one being extremely inaccurate and nine being extremely accurate. The second survey consisted of seven questions asking participants to rate (on nine-point scales) how comfortable they were conversing with the bot (1 = extremely uncomfortable to 9 = extremely comfortable), and perceptions of the bot's humanness (1 = extremely inhuman like to 9 = extremely humanlike), conversational skill (1 = extremely unskilled to 9 = extremely skilled), thoughtfulness (1 = extremely unthoughtful to 9 = extremely thoughtful), politeness (1 = extremely impolite to 9 = extremely polite), responsiveness

(1 = extremely unresponsive to 9 = extremely responsive), and how engaging he was (1 = extremely unengaging to 9 = extremely engaging).

### 8.3. Procedure

This experiment was conducted in a lab room rather than on-line. When participants arrived for a session they were told they would chat, via a computer, with an artificial agent named Pat. They were told to become acquainted with Pat, but to limit their remarks to non-offensive topics, such as sports, music, classes, movies, travel and so on. Participants were randomly assigned to either the name or no name condition and were not aware that they were in any particular condition. In the name condition, the experiment would input the participant's name and gender before the chat began. Once the chat began the experimenter left the room. This version of the bot had a tendency to crash and participants were informed that this was the case. They were also instructed to click on the bot icon to restart the chat if it crashed. After 10 min, the experimenter returned and participants were asked to complete the two questionnaires described above. When finished, participants were debriefed and thanked for their participation.

## 9. Results and discussion

The chats were examined to ensure that the name manipulation was successful. The mean number of times the bot used the participant's name was 4.8 ($s = 1.64$) in the name condition. A participant's name was never used in the no name condition.

We first conducted a factor analysis of the six rating scales in order to determine the underlying dimensions for evaluating the bot. Following the recommendations of Russell (2002), a principle axis extraction with varimax rotation was used. Two factors emerged with eigenvalues greater than one that together accounted for 74% of the variance. The first factor involved perceptions of the bot's conversation skill and was defined by three variables (with their corresponding loadings on this factor): how human was the bot (0.81), how skilled was the bot (0.93), and how engaging was the bot (0.61). The second factor involved perceptions of the bot's pleasantness and was defined by three factors: how polite was the bot (0.656), how thoughtful was the bot (0.841), how responsive was the bot (0.501).

Two factor scores were created (using the regression method option)[2] and analyzed with an analysis of variance with a single independent variable (name vs. no name). For the first factor, there was a significant name effect, $F(1,47) = 4.11$, $p < 0.05$; participants perceived the bot to be more skilled when he used their name ($M = 0.274$) than when he did not use their name (−0.263). Although the bot was perceived to be more pleasant when he used participant's names (0.18) than when he did not (−0.173), the effect for the second factor was not significant, $F(1,47) = 1.93$, $p < 0.20$ (see Table 2).

We were also interested in perceptions of the bot on the five-factor model. Hence, responses to the 100 traits from the five-factor model were combined to create composite scores on the five factors. The bot was perceived more positively on each of these five traits when he used the participant's first name relative to when he did not. However,

[2] With a regression method the scores produced have mean of 0 and a variance equal to the squared multiple correlation between the estimated factor scores and the true factor values.

Table 2
Experiment 2: Perceptions of bot as a function of using participant's first name

| | First name used | | | |
|---|---|---|---|---|
| | Yes | No | *F* | *p* |
| *Factor 1 conversational skill* | | | 4.11 | <0.05 |
| Skilled | 4.62 | 3.52 | 5.58 | 0.03 |
| Human | 4.54 | 3.68 | 2.17 | 0.15 |
| Engaging | 5.79 | 4.92 | 2.41 | 0.13 |
| *Factor 2 pleasantness* | | | 1.93 | <0.20 |
| Thoughtful | 5.08 | 4.12 | 3.86 | 0.055 |
| Polite | 6.12 | 5.48 | 1.15 | ns |
| Responsive | 6.42 | 5.96 | <1 | ns |

none of these differences were significant (all $p > 0.05$). Unlike the more specific dimensions described above, perceptions of the bot's personality did not vary significantly as a function of the name manipulation. We postpone discussion of these results until later.

Overall, then, the results of this demonstrate that a rather subtle manipulation – whether the bot occasionally used the participant's first name – can have a generally positive effect on perceptions of a bot. Although the differences were not all significant (due in part to relatively high rating variance) they were consistently all in the same direction: when the bot used the participants first name he was perceived more positively than when he did not use their names. And this effect was largest and significant for the overall conversational skill factor.

## 10. Experiment 3

In this experiment, we examined the impact of a different feature of conversational style – response latency – on perceptions of the bot. We chose this variable for two reasons. First, research on human–computer interaction has demonstrated that response latency is an important variable that influences user satisfaction (Card, Moran, & Newell, 1983; Schneiderman, 1998). Second, research on human communication suggests that response latency is an important dimension of conversation style, a variable that both reflects personality traits, such as extraversion (Ramsey, 1966) and influences perceptions of the speaker in terms of competence (Scherer, 1979). Hence, response latency should play a role in perceptions of a bot's personality.

## 11. Method

### 11.1. Participants

Participants ($N = 93$) were students enrolled in introductory psychology classes at Ball State University who participated for partial course credit.

### 11.2. Procedure and materials

This study was conducted on-line. The procedure and materials (post-chat questionnaire) were identical to that used in Experiment 1 with the following exception. One-half

of the participants interacted with a bot programmed to respond quickly (1 s delay) to the users remarks. The remainder chatted with a bot programmed to respond more slowly (10 s delay) to the users remark.

## 12. Results and discussion

Similar to Experiment 1, some of the students had difficulty accessing the bot and failed to provide enough information on the questionnaire to calculate composite measures. Due to these issues analyses were limited to the 64 participants who provided complete data.

Participants ratings of the bot's responsiveness served as a manipulation check. The delay manipulation was successful and participants in the short delay condition rated the bot as more responsive ($M = 7.27$) than did participants in the long delay condition ($M = 4.61$), $t(62) = 2.32$, $p < 0.05$.

Composite measures of the five factors were created and analyzed as a function of delay. These data are presented in Table 3. The delay manipulation had a significant effect on two factors. Participants perceived the bot as higher on surgency (106.35 vs. 95.73, $t(50) = 2.32$, $p < 0.05$) and conscientiousness (111.5 vs. 102, $t(54) = 2.83$, $p < 0.05$) when he responded quickly than when he responded more slowly. No other effects were significant.

## 13. General discussion

The major theme to emerge from the present set of studies is that conversation bots clearly can be viewed as having human-like personalities. First, there was significant agreement between participants in their perceptions of the bot's major personality traits. This is important because observer consensus is a necessary (though not sufficient) criteria for demonstrating the existence of personality (Kenrick & Funder, 1988). Hence, participants' ratings did not simply reflect their own idiosyncratic views of the bot. Rather, there was something about the bot – i.e., its "personality" – that influenced these ratings and produced the consensus observed in this research. Note in this regard, that the traits from the five-factor model (or Big 5) appear to be relevant for perceiving bots. Future research on human–bot interaction should probably continue to use these traits.

Table 3
Experiment 3: Bot perceptions as a function of response delay

| | Delay | | | | |
|---|---|---|---|---|---|
| | Short | | Long | | |
| Trait: | Mean | SD | Mean | SD | $t$ |
| Conscientiousness | 111.5 | 12.0 | 102.0 | 13.09 | 2.83* |
| Neuroticism | 90.38 | 17.92 | 88.48 | 18.4 | 0.39 |
| Surgency | 106.35 | 14.1 | 95.74 | 18.61 | 2.32* |
| Agreeableness | 116.87 | 17.96 | 109.75 | 18.28 | 1.51 |
| Openness to experience | 103.96 | 17.87 | 97.6 | 18.35 | 1.26 |

* $p < 0.05$.

Second, perceptions of the bot were influenced by communication variables that have been demonstrated to influence perceptions of human communicators. Specifically, a relatively high degree of familiarity on the part of the bot – and hence a positively polite style – resulted in perceptions of a more skilled conversationalist. And a bot who responded relatively quickly was perceived as more conscientious and extraverted than a bot who responded more slowly. Finally, similar to human interactants, participants preferred not to chat (as evidenced by shorter chat times) with a bot they perceived as being high in neuroticism.

The present results suggest that artificial agents have the potential to work well as points of contact on web sites. One issue for web site bot developers is whether a bot should mimic human qualities, or whether it should convey a distinctly non-human, robot-like persona. This is a fundamental issue for human–computer and human–robot interaction. Can a bot be made to be too human? Research on human–robot interaction certainly suggests this as a possibility (e.g., Dautenhahn et al., 2002). This is why the present research is important – there is a clear need to uncover the parameters of successful human–computer communication.

Our results suggest that this humanness issue depends, in part, on the specific human qualities one chooses to include in the bot. A more personable bot – one that uses its' interlocutor's first name – will be perceived more positively, clearly a plus for the web site. However, there are clear cultural limits to such preferences due to cultural differences in the preference for positive politeness (Holtgraves, 2002). Hence, people from cultures that prefer a more formal interaction style may be put-off by a positively polite bot that uses the users first name. Similarly, a bot that responds quickly may be perceived more positively than a slowly-responding bot. But here too, culture matters and will probably limit this effect.

There are additional, related issues that could be pursued in this line of research. For example, do people prefer interacting with a ''male'', ''female'', or androgenous bot? What exactly are the limits to how personable a bot should be? What human qualities work best in a bot? And so on. In addition to web site use, conversation bots may prove to be a useful tool for investigating various aspects of social interaction in general, and conversational interaction in particular. In fact, the primary motivation for developing the bot used in the present research was to create a tool for examining online language comprehension (see Holtgraves & Han, in press). In these studies, participants converse with a bot but believe that they are conversing with another human. During these chats, participants periodically perform a secondary task (e.g., lexical decision task). In addition, the time taken to comprehend the bot's utterances is recorded. These (and other) measures allow for the examination of the cognitive processes involved in language comprehension for people who are actually engaged in an interaction (rather than reading or listening to the remarks of others).

It seems quite likely that such a procedure could be easily adapted for use in studying many features of social interaction. For example, a bot could be programmed with different personalities, and the impact of those personalities on the content and course of interactions could be examined. Such a procedure would be particularly useful for studying gender effects. Participants could chat with a bot that was described as either male or female. Then, the impact of assumed gender on the content and course of an interaction could be examined in a very controlled manner.

There is no doubt that the use of artificial agents will continue to increase, a phenomenon that is worthy of psychological investigation. Psychological research can both inform

the development of these natural language interfaces as well as benefit from their use in the study of social interaction processes.

## Acknowledgements

This research was supported by grants from the National Science Foundation (SBR-0131877) and an Icommunication Grant awarded to Ball State University. The assistance of Frank Carr in conducting Experiment 2 is greatly appreciated.

## References

ALICE. Artificial Intelligence Foundation. http://www.alicebot.org/ (retrieved 14.02.06).

Botspot. http://www.botspot.com (retrieved 14.02.06).

Breazeal, C. (2002). *Designing sociable robots*. Cambridge, MA: MIT Press.

Brennan, S. E. (1997). The grounding problem in conversations with and through computers. In S. R. Fussell & R. J. Kreuz (Eds.), *Social and cognitive approaches to interpersonal communication* (pp. 201–228). Mahwah, NJ: Erblaum.

Brown, P., & Levinson, S. (1987). *Politeness: Some universals in language usage*. Cambridge, UK: Cambridge University Press.

Card, S. K., Moran, T. P., & Newell, A. (1983). *The psychology of human–computer interaction*. Hillsdale, NJ: Erlbaum.

Cohen, P. R., & Levesque, H. J. (1990). Rational interaction as the basis for communication. In P. R. Cohen, J. Morgan, & M. E. Pollack (Eds.), *Intentions in communication* (pp. 221–256). Cambridge, MA: MIT Press.

Colby, K. A. (1972). Artificial paranoia. *Artificial Intelligence, 2*, 1–26.

Dautenhahn, K., Ogden, B., & Quick, T. (2002). From embodied to socially embedded agents: implications for interaction aware robots. *Cognitive Systems Research, 3*, 397–428.

Digman, J. M. (1990). Personality structure: emergence of the five-factor model. *Annual Review of Psychology, 41*, 417–440.

Goldberg, L. R. (1992). The development of markers for the big-five factor structure. *Psychological Assessment, 4*, 26–42.

Holtgraves, T. (2002). *Language as social action: Social psychology and language use*. Mahwah, NJ: Erlbaum.

Holtgraves, T., Han, T.L. (in press). A procedure for studying on-line conversational processing using a chat bot. *Behavior Research Methods*.

Hovy, E. H. (1988). *Generating natural language under pragmatic constraints*. Hillsdale, NJ: Erlbaum.

Isbister, K., & Nass, C. (2000). Consistency of personality in interactive characters: verbal cues, non-verbal cues, and user characteristics. *International Journal of Human-Computer Studies, 53*, 251–267.

Kashdan, T. B. (2002). Social anxiety dimensions, neuroticism, and the contours of positive psychological functioning. *Cognitive Therapy and Research, 26*, 789–810.

Kenrick, D. T., & Funder, D. C. (1988). Profiting from controversy: lessons from the person-situation controversy. *American Psychologist, 43*.

Lee, K. M., & Nass, C. (2004). The multiple source effect and synthesized speech: doubly disembodied language as a conceptual framework. *Human Communication Research, 30*, 182–207.

Nass, C., Moon, Y., & Green, N. (1997). Are machines gender neutral? Gender-stereotypic responses to computers with voices. *Journal of Applied Social Psychology, 27*, 864–876.

Ramsey, R. (1966). Personality and speech. *Journal of Personality and Social Psychology, 4*, 116–118.

Reeves, B., & Nass, C. (1996). *The media equation*. Stanford, CS: CSLI Publications.

Reichman, R. (1985). *Getting computers to talk like you and me: Discourse context, focus, and semantics*. Cambridge, MA: MIT Press.

Rheingold, H. (1985). *Tools for thought: The people and ideas behind the next computer revolution*. New York: Simon & Schuster.

Rosenthal, R., & Rosnow, R. (1991). *Essentials of behavioral research: Methods and data analysis* (second ed.). New York: McGraw-Hill.

Russell, D. W. (2002). In search of underlying dimensions: the use (and abuse) of factor analysis. *Personality and Social Psychology Bulletin, 28*, 1629–1646.

Schank, R. C. (1985). *The cognitive computer: On language, learning, and artificial intelligence*. Reading, MA: Addison-Wesley.

Scherer, K. (1979). Personality markers in speech. In K. Scherer & H. Giles (Eds.), *Social markers in speech* (pp. 147–209). Cambridge, UK: Cambridge University Press.

Schneiderman, B. (1998). *Designing the user interface: Strategies for effective human–computer interaction* (third ed.). Reading, MA: Affison-Wesley.

Sproull, L., Subramani, M., Kiesler, S., Walker, J. H., & Waters, K. (1996). When the interface is a face. *Human-Computer Interaction, 11*, 97–124.

Weizenbaum, J. (1966). Eliza – a computer program for the study of natural language communication between man and machine. *Communication of the ACM, 9*, 36–45.