Available online at www.sciencedirect.com

ScienceDirect

Computers in Human Behavior 24 (2008) 1494–1509

Computers in Human Behavior

www.elsevier.com/locate/comphumbeh

# How social is social responses to computers? The function of the degree of anthropomorphism in computer representations

Li Gong *

*School of Communication, Ohio State University, 3149 Derby Hall, 154 N. Oval Mall, Columbus, OH 43210, United States*

Available online 10 July 2007

**Abstract**

Testing the assumption that more anthropomorphic (human-like) computer representations elicit more social responses from people, a between-participants experiment ($N = 168$) manipulated 12 computer agents to represent four levels of anthropomorphism: low, medium, high, and real human images. Social responses were assessed with users' social judgment and homophily perception of the agents, conformity in a choice dilemma task, and competency and trustworthiness ratings of the agents. Linear polynomial trend analyses revealed significant linear trends for almost all the measures. As the agent became more anthropomorphic to being human, it received more social responses from users.



*Keywords:* Anthropomorphism; Computer representations; Agents; Human-like; Social responses to computers

## 1. Introduction

An often assumed psychological process in people's interaction with computers and particularly representations on computers such as agents is that the more human-like the computer representation is, the more social people's responses are. Anthropomorphism, enabling computers with human-like characteristics and capabilities, has been a

* Corresponding author. Tel.: +1 614 247 8435; fax: +1 614 292 2055.
*E-mail address:* ligong@stanfordalumni.org

0747-5632/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.chb.2007.05.007

driving philosophy behind the advances of computer technology (anthropomorphism in this paper refers to the technological efforts of imbuing computers with human characteristics and capabilities). Anthropomorphic efforts in computing range from artificial intelligence for the underlying algorithm (Turing, 1950) to humanized interface representations such as faces (Gong & Nass, 2007; Sproull, Subramani, Kiesler, Walker, & Waters, 1996) and voices (Gong & Lai, 2003; Nass & Brave, 2005).

The body of media equation research (Nass & Moon, 2000; Reeves & Nass, 1996) through numerous studies demonstrated that users apply social rules to their interaction with computers. Computers in most of the media equation studies only had text output rather than highly or obviously anthropomorphic features such as computer-synthesized faces or voices. Although computers with text output nonetheless elicited social responses, Nass and Moon (2000) further posited a possible hypothesis that "the more computers present characteristics that are associated with humans, the more likely they are to elicit social behavior" (p. 97). In fact, a linear relationship is often assumed between the degree of anthropomorphism in computers and socialness in users' responses. This assumption underlies many sophisticated anthropomorphic endeavors such as facial animation, speech synthesis and recognition, natural language interaction, and artificial intelligence. Computer interface agents are a quintessential example of anthropomorphic computer representation because they are often visually embodied with human-featured synthetic faces and sometimes coupled with synthetic speech and supported by conversational capability and artificial intelligence (Cassell, Sullivan, Prevost, & Churchill, 2000; Maes, 1994).

The motives behind agent endeavors have two layers of assumptions: (1) the more anthropomorphic a computer representation is, the more socially users respond to it; (2) and as a result, the more effective the human–computer interaction is. The debate on agents and anthropomorphism has mainly centered around the second assumption regarding effectiveness. Dehn and van Mulken (2000) provided an extensive review of the research on human–agent interaction in this regard and painted a picture of inconsistent results regarding the effects of agents. They elegantly pointed out a list of methodological issues accounting for the inconsistency across studies. One prominent issue is the large variability in embodiments of the agents which may range from simplistic 2D caricature drawings to 3D head models.

Besides various methodological problems raised by Dehn and van Mulken, a matching hypothesis has recently emerged in the literature. It proposes to congruently match the use or design of agents and the degree of anthropomorphism with the nature and goal of the human–computer interaction to achieve effectiveness and optimal performance (Burgoon et al., 2000; Dehn & van Mulken, 2000; Goetz, Kiesler, & Powers, 2003; Gong & Nass, 2007). For example, an interaction task which does not involve a social orientation would not call for high anthropomorphism.

What is still left unverified with inconclusive evidence, however, is the aforementioned first assumption that more anthropomorphic computer representations lead to more social responses. This relationship is independent of the second assumption about effectiveness because social responses may or may not be conducive to the effectiveness of a specific context of human–computer interaction (Burgoon et al., 2000; Goetz et al., 2003). The anthropomorphism–social responses relationship is a basic theoretical proposition to examine because it underlies many research efforts and people's interaction with computers in all contexts.

A few empirical studies either directly tested or partially concerned the assumed linear relationship between anthropomorphism and social responses. The results are far from being consistent and conclusive. A review of these studies reveals a few key conceptual and methodological issues: (1) conceptual and operational confounding of the degree of anthropomorphism with modalities of computer representations; (2) lack of control for factors such as physical attractiveness which are salient in perceiving visual representations; (3) experimental tasks which do not unequivocally tap into social responses; and (4) unclear conceptualization of the plain-text computer interface as a control condition.

### *1.1. Testing anthropomorphism and social responses*

Sproull et al. (1996) conducted one of the first experimental tests on the notion of anthropomorphism and social responses. They anthropomorphized the computer interface with a talking face, which was a video-captured real human face texture-mapped on a geometric wire-frame and animated to speak with DEC-Talk computer-synthesized speech. DEC-Talk speech was a traditional type of formant-based synthetic speech and sounded mechanistic and obviously artificial, whereas the face clearly looked of a real human. Participants interacted with the talking face or plain text output on a computer, answering psychological scales and discussing career interests. Participants gave more socially desirable responses to the talking face than to the text interface, suggesting more social responses to the face. But in the post-interaction questionnaire, they rated the text interface more positively than the talking face in social evaluation and sociability.

Adding a condition with only the DEC-Talk voice and a condition involving face-to-face (FtF) interaction with a human confederate, Kiesler, Sproull, and Waters (1996) compared the same talking face and text output in terms of people's cooperativeness in a prisoner's dilemma task and ensuing social judgment. Again, the talking face received lower social evaluation than the voice alone or the text interface. But, the FtF human condition received the highest social evaluation rating. The same pattern of results was found for cooperativeness. The mechanistic-sounding DEC-Talk voice cannot explain the low social evaluation of the talking face, an explanation suggested by Dehn and van Mulken (2000), because the voice alone received higher evaluation than the talking face. The culprit dampening the talking face is most likely to be the mismatch between the real human face and the synthetic voice. Gong and Nass (2007) found that mismatching an agent's face and voice in the real human versus computer-synthesized dimension created inconsistency and caused negative effects on trust, disclosure, perception, and judgment.

The variation of the talking face, the voice, and the text inherently confounded the degree of anthropomorphism with modalities of the computer representation. Multimodality then invites the potential issue of inconsistency. This same issue of confounding was inherent in the design of the study by Burgoon et al. (2000).

Burgoon et al. (2000) set out to verify the posited linear relationship between modality richness and the degree of anthropomorphism in computer agents and credibility ratings and influence. They systematically varied five versions of computer agent representations: text-only; text and a synthesized voice; text, the voice, and a still image of a computer-synthesized face; the face animated and talking with the voice; and the talking face with text display. They also added two FtF human confederate conditions which varied in interactivity with one condition allowing the confederate to respond to the participants' remarks. Persuasive influence was assessed with the Desert Survival Task, which was followed by a host of credibility ratings. The main results were that the FtF human conditions received higher sociability ratings than the computer ones. Among the computer interfaces, an approaching significance linear trend featured the text interface with the highest rating in sociability. Through examining the means, the text interface had similar ratings as the FtF human conditions. The text, voice and still image condition had the lowest ratings, which is again interpretable from a modality-matching perspective because the still image was not synched with the voice. Puzzlingly, the FtF human conditions were less influential than the computer conditions. But the mean examination suggests that the low influence mainly existed in the FtF condition allowing interaction between the participant and the confederate. Thus, it is unclear whether the low influence was due to the interactivity or the possibility that computers may receive more expertise attribution for items involved in the Desert Survival Task such as rope, matches, and knife.

The results of these three studies showed the puzzling superiority of the text interface, relative to synthetic faces and/or voices. They also suggested the need to question the appeal of the particular computer-synthesized faces and voices used. Furthermore, the variation of modalities to realize the level of anthropomorphism invites confounding and the potential issue of mismatching between modalities. From the modality perspective, it is also problematic to construe the text interface as a low-anthropomorphism operationalization because it does not operate in the same mode as face or voice. Nonetheless, these three studies lent some support to the relationship between the degree of anthropomorphism and social responses because even the poorly rated talking face received more socially desirable responding than the text interface (Sproull et al., 1996) and FtF human interaction received the highest social judgment (Burgoon et al., 2000; Kiesler et al., 1996).

Within the modality of facial image, Nowak (2004) compared people's ratings of social attraction and credibility of a more anthropomorphic synthetic face and a less anthropomorphic one. A no-image control condition was also included. All conditions presented the verbal content through the recorded voice files of a real woman. Participants exchanged self-introduction and Web scavenger skills with the computer representation, which was framed as either an agent or an avatar (this manipulation did not cause any differences). Contrary to Nowak's hypotheses, the linear pattern in the results showed that the simplistic less anthropomorphic face received the highest ratings of social attraction and credibility whereas the more anthropomorphic face ranking the lowest.

Using the same experimental manipulation, Nowak and Biocca (2003) found similar results: the less anthropomorphic face received the highest ratings of telepresence, copresence, and social presence. Nowak and Biocca reported open-ended comments from some participants that the more anthropomorphic face was "not very attractive" and "funny looking" (p. 491). They proposed expectation as an explanation in that a relatively more realistic synthetic face elicits higher expectation. Another explanation as evident in the participants' comments was judgment of physical attractiveness. The two facial images used in their study were included in the pilot test of 85 computer-synthesized images of various characters in the research presented here (details are in Section 2). Their less anthropomorphic image positioned at the 82nd percentile of the attractiveness rating, whereas their more anthropomorphic image was at the 75th percentile.

When computer characters were perceived as more attractive, social judgment of them was more positive. Lee and Nass (2002) found that animated characters elicited more positive social presence judgment and were rated more attractive and trustworthy than stick figures and text boxes in a computer-based social influence task using choice dilemma scenarios. Their animated characters overall appeared more anthropomorphic and appealing than the stick figures. Although Lee and Nass did not aim to test the degree of anthropomorphism, their results lent some support to the notion of anthropomorphism and social responses and indicated the potency of physical attractiveness.

Employing a within-participants design, Goetz et al. (2003) tested people's assignment of 12 variations of 2D robot heads to various jobs. Confirming their hypothesis that the more anthropomorphic appearance elicits more social attribution, more human-like robots were assigned to jobs involving more social interaction and more machine-like robots were assigned to jobs involving less social interaction.

### *1.2. An integrative critique and hypotheses*

The extant studies testing the assumed relationship between the degree of anthropomorphism in computer representations and people's social responses provided inconsistent and inconclusive evidence. Conceptual and methodological problems need to be solved before proper tests can be conducted.

First, the modality of computer representations needs to be conceptually and operationally separated from the degree of anthropomorphism. Most of the studies that have been reviewed confounded modalities with anthropomorphism, making it difficult to tease apart the effects of anthropomorphism. Computer-synthesized faces can vary from very low human-likeness to very high human-likeness, so can computer-synthesized voices (Gong & Lai, 2003) and wording styles in text. Hence, anthropomorphism can vary in each modality as well as in combinations of modalities. In the latter case, problems of mismatching and incongruence are possible as in the talking face used by Sproull et al. (1996) and Kiesler et al. (1996) and the voice with still image used by Burgoon et al. (2000). To allow a pure test of the degree of anthropomorphism, it is the best and parsimonious to limit to one type of modality. The research presented in this paper limits to still facial images because faces are a prominent feature of computer representations. Confinement within facial images will help improve experimental internal validity in that degree of anthropomorphism rather than modality is the concept under study.

Second, another issue related to modality is the choice of communication medium in operationalizing the ideal prototype of anthropomorphism: humans. It was operationalized as FtF human interaction in Burgoon et al. (2000) and Kiesler et al. (1996). However, the variation from computer-based anthropomorphic representations to FtF human interaction changes two things at once: the communication partner being human-like versus being human; the communication medium being computer versus FtF. This research limits only to the computer medium, presenting facial images of real people as well as anthropomorphic facial images all on computers.

Third, other potentially confounding factors for perceiving and judging facial images need to be controlled for, mainly physical attractiveness and expressions. Although the reviewed studies did not appear to have issues with expressions, attractiveness was not reported to be controlled for. Attractiveness seemed to cause confounding in the studies of Nowak (2004) and Nowak and Biocca (2003).

Fourth, the experimental task for capturing social responses needs to unequivocally point to the social direction rather than possibly be contaminated by other expertise concerns. Desert Survival Task (Burgoon et al., 2000) and Web scavenger skills (Nowak, 2004), for example, do not appear to live up to this requirement. As Burgoon et al.

(2000) pointed out, computers may be assigned more expertise to matters associated with technology. The choice dilemma scenarios used to assess computer-based social influence in Lee and Nass (2002) and Lee (2004) seem to be a more suitable experimental task to embody the test of social responses because all the scenarios involve purely social matters (marriage, family, career choice, personal investment, and political campaign). A more human-like entity should carry more weight in exerting influence in these social matters than a less human-like one.

Fifth, one consistent puzzling result in the extant studies is the superior judgment of the text interface compared to more anthropomorphic representations such as faces and voices (Burgoon et al., 2000; Kiesler et al., 1996; Sproull et al., 1996). The notion of default image posited that the lack of image does not mean that people do not at least unconsciously project the image of an entity according to their prototypical experiences (Nowak & Biocca, 2003). Burgoon et al. (2000) proposed a similar discussion that less cues do not equate to crippled perception or inferior judgment, leveraging Walther (1996) concept of hyperpersonal computer-mediated communication. The basic idea is that people may automatically enrich their perception by adding readily available prototypical information from one's memory. These contemplations suggest that the text interface operates under a different frame of comparison from interfaces with facial images. Simply, text presents a different modality from faces. Thus, a text interface is not an appropriate comparison to a face-enhanced interface for the purpose of testing the degree of anthropomorphism.

To sum, a stringent test of anthropomorphism should control the modality factor. This research focused on varying the level of anthropomorphism in facial images as computer representations. All facial images would be in the computer medium. The facial images would be subject to a pilot test of physical attractiveness. And, social influence through choice dilemma scenarios would constitute a suitable test bed for social responses.

A series of hypotheses were formulated to test the relationship between the degree of anthropomorphism and people's social responses. First, as all previous studies assessed social judgment of computer representations, it was hypothesized that:

H1: The more anthropomorphic a computer representation is, the more positive social judgment it will receive from people.

Another concept that has not been explored by prior studies on anthropomorphism is the perception of homophily. All other things being equal, being more human-like is expected to elicit greater perception of similarity from people.

H2: The more anthropomorphic a computer representation is, the greater homophily perception it will receive from people.

In purely social matters, a more human-like computer representation is hypothesized to be more influential than a less human-like one.

H3: The more anthropomorphic a computer representation is, the greater social influence it will exert on people.

In the literature on influence and persuasion, credibility is considered a critical factor (Burgoon et al., 2000). Credibility includes the components of competence/expertise, trustworthiness, and dynamism or attractiveness (Berlo, Lemert, & Mertz, 1970; McCroskey, Hamilton, & Weiner, 1974; Ohanian, 1990). Because attractiveness would be controlled for, this research focused on ratings of competence and trustworthiness. The ratings of competence and trustworthiness were also expected to vary with the degree of anthropomorphism.

H4: The more anthropomorphic a computer representation is, the more competent it will be perceived to be by people.
H5: The more anthropomorphic a computer representation is, the more trustworthy it will be perceived to be by people.

Although text interface has been construed as not comparable to facial representations, it was included in this research because it has been included in almost all the studies in the literature. No hypotheses were made with respect to it. Its inclusion was to provide another test of how text interface would compare to interfaces with facial representations. To note again, the text interface was not conceptualized as a control condition or a variation in the anthropomorphism degree.

## 2. Method

A between-participants experiment with a one-way design was conducted. People interacted with facial images varying in the degree of anthropomorphism or with text output in the computer interface. A key issue that has not received adequate treatment in the literature is how to operationalize the degree of anthropomorphism. A facial representation considered highly anthropomorphic in one study might be considered less anthropomorphic in another study. To bear an externally valid connection with the phenomenon of computer-synthesized facial images of various kinds of entities, this research based the variations of anthropomorphism on a large collection of images.

### *2.1. Collection and pilot test of facial images*

The images of 610 computer-synthesized characters which all contained faces were collected from online digital galleries, computer games, computer applications, and research papers. None of the images was of a real human. They were all computer-synthesized. Then, 85 images were randomly sampled from this large pool. In a pilot study, 273 undergraduate students (120 men, 153 women) enrolled in two large communication courses of a large US Midwestern public university rated all the 85 images in a character perception survey. The participants were awarded with extra course credit. On a Web site created for the pilot study, the 85 images were presented one per page in a random order. The participants judged each image in terms of the degree of human-likeness (*not human-like–human-like*) and physical attractiveness (*ugly–beautiful*) on 7-point semantic differential scales. The questionnaire also contained other perception measures of digital characters for other research purposes.

The ratings of human-likeness and attractiveness were averaged across all participants. Then, the 85 images were broken down into three approximately equal-sized groups by their percentiles in the human-likeness rating, corresponding to low, medium, and high levels of anthropomorphism. Three ranks rather than a high versus low dichotomy were used to allow a more fine-grained operationalization of the degree of anthropomorphism. To reduce the effects of idiosyncrasy of any given facial images, three images were randomly selected from each of the three levels of anthropomorphism to instantiate each level except with the constraint of comparativeness in attractiveness and expressions.

The selected images were controlled to be comparable in attractiveness and expressions. All the facial images selected fell in the inter-quartile range (25th to 75th percentiles) of the attractiveness rating. They all had forward-looking facial orientation and relatively neutral facial expressions.

Hence, the three degrees of anthropomorphism were instantiated by nine computer-synthesized facial images (see Fig. 1). Two of the three low-anthropomorphism images were humanoid robot characters. They were humanoid, i.e., human-like. Being human-like is conceptually different from being perceived as human. These two humanoid robot characters and the simplistic face as used by Nowak (2004) and Nowak and Biocca (2003) all belonged to the low level of human-likeness.

Consistent with previous studies (Burgoon et al., 2000; Kiesler et al., 1996), human representations were included as the ideal comparison of anthropomorphism. To keep the

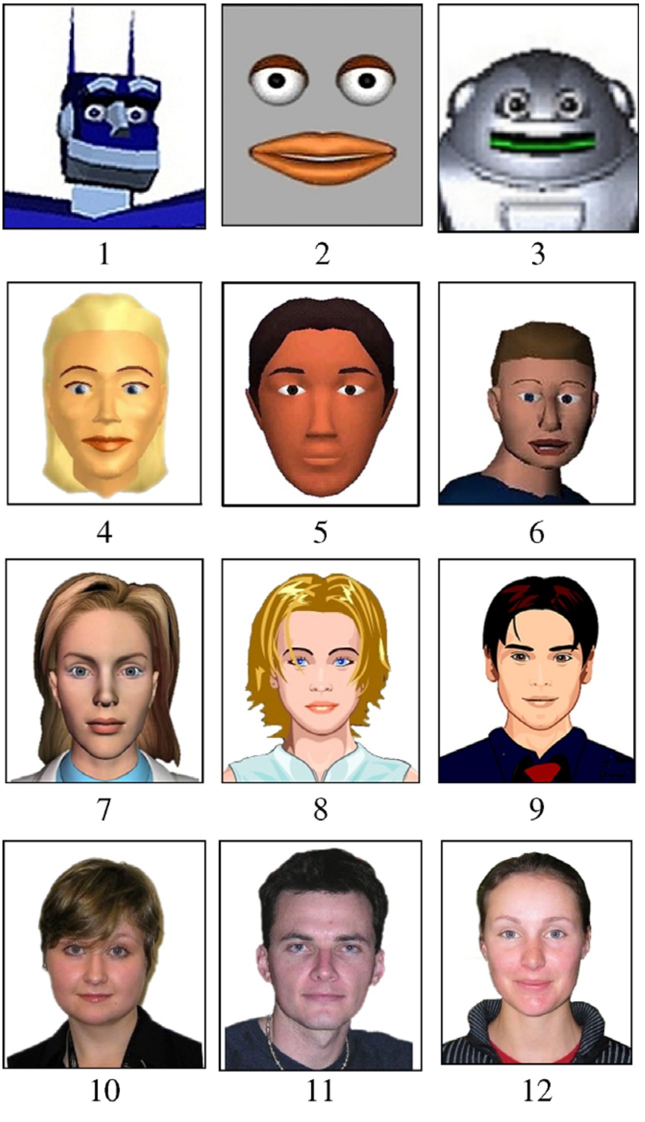


Fig. 1. The facial images used as computer representations in the study.

computer medium constant, images of real humans were presented on the computer. Three human facial images were randomly selected from another pool of facial images of 80 people. These 80 human facial images were rated in a host of perceptual dimensions including physical attractiveness in another perception study. The selected three facial images also fell in the inter-quartile range of attractiveness rating among the 80 human faces, had forward-looking orientation, and presented relatively neutral facial expressions. The three human facial images are also presented in Fig. 1.

### 2.2. Participants

Participants in the main experiment were a different sample of 168 undergraduate students (80 males, 88 females) recruited from large communication courses in the same university. Thirty-six participants were randomly assigned to each of the four levels of anthropomorphism including real human images, with 12 participants per image. Twenty-four participants were assigned to the text-output condition. Gender of the participants was balanced for each image and condition. The participants were rewarded with extra course credit.

### 2.3. Procedure and the experimental instrument

The social influence instrument of five choice dilemma scenarios used by Lee and Nass (2002) was employed as the interaction task between the participants and the computer. These dilemma scenarios of social matters presented two competing choices: one is more risky but suggests more rewards and the other one is safer but brings less rewards. These five scenarios concerned marriage, family, career choice, personal finance, and political campaign (see Lee & Nass, 2002 for details). The facial image was framed as an agent in the choice dilemma scenarios (the text output version was referred to as the system). The task was conducted on a computer with a 17 in. flat panel monitor.

The participants first read each scenario on the computer screen. On the next screen, the facial image of the agent was displayed on the left of the screen presenting its proposed decision for the dilemma. The agent's decision was presented in a text box to the right of the agent. The text box had an arrow pointing to the agent symbolizing the agent as the source of the proposed decision (Fig. 2 presents a sample screenshot). In the text condition, the text box was positioned in the center of the screen without any facial image. One second after the agent (or the system in the text condition) presenting its decision, the same computer screen showed six options for the participants to choose from to indicate their decision. Following Lee and Nass (2002), the six options ranged from "definitely should do", "should do", and "probably should do" one of the two choices. One example is "going to the conservatory of music" versus "going to the medical school". The participants' responses were coded as 1–6 with 1 indicating maximum disagreement and 6 indicating maximum agreement. Then, same as Lee and Nass (2002), the participants typed their reason in a text box on the next screen, with the agent (or the system) presenting its reason on the last screen of each scenario.

After all the five choice dilemma scenarios, the participants filled out a questionnaire on the computer. The questionnaire contained measures on social judgment, homophily, competence, and trustworthiness. At last, the participants were debriefed and thanked.

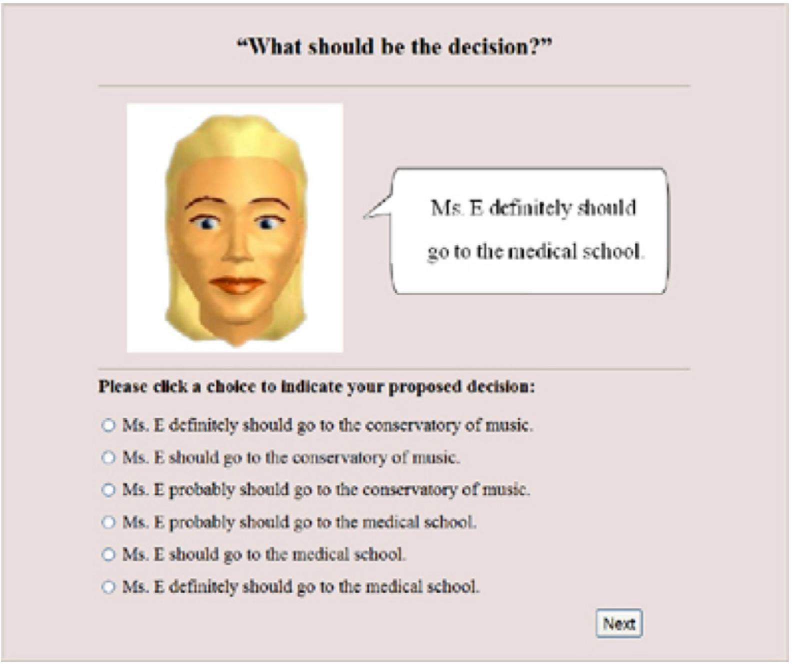


Fig. 2. A sample screenshot of one of the facial images in one of the choice dilemma scenarios.

### 2.4. Measures

The social influence measure in the choice dilemma scenarios was computed following Lee and Nass (2002) $z$-transformation procedure to control for the participants' predispositional risk-taking tendency (for details on the procedure and rationale, see Lee & Nass, 2002). Then, the $z$ scores of the responses to the five scenarios were averaged to form the social influence measure ($\alpha = .71$). Higher $z$ scores indicate greater social influence.

Social judgment of the computer representation was measured with the four 7-point semantic differential items developed by Short, Williams, and Christie (1976) in their social presence research: *impersonal–personal*, *sociable–unsociable* (reverse coded), *insensitive–sensitive*, *warm–cold* (reverse coded) ($\alpha = .86$). The responses to these four items were averaged, with higher scores indicating more positive social judgment. Short et al.'s scale was widely used as judgment of socialness in previous research (e.g., Lee & Nass, 2002; Rice & Love, 1987). Other social judgment measures such as social attractiveness used by Sproull et al. (1996) and sociability used by Burgoon et al. (2000) appear to be highly similar to these four items.

Perceived homophily of the computer representation was measured with the four applicable items from McCroskey, Richmond, and Daly's (1975) homophily scale: *is like me–is unlike me* (reverse coded), *is different from me–is similar to me*, *thinks like me–does not think like me* (reverse coded), and *does not behave like me–behaves like me* ($\alpha = .93$). The responses to these four 7-point items were averaged, with higher scores indicating greater homophily.

Competence rating of the computer representation used applicable items from competency subscales in credibility research (Berlo et al., 1970; McCroskey et al., 1974; Ohanian, 1990). It included six pairs of 7-point semantic differential scales: *unintelligent–intelligent*, *knowledgeable–unknowledgeable* (reverse coded), *incompetent–competent*, *uninformed–informed*, *inexpert–expert*, *experienced–inexperienced* (reverse coded) ($\alpha = .90$). Responses to these six items were averaged, with higher scores indicating higher competence ratings.

Trustworthiness rating of the computer representation used nine applicable items from Wheeless and Grotz (1977) Individualized Trust Scale: *trustworthy–untrustworthy* (reverse coded), *unreliable–reliable*, *insincere–sincere*, *honest–dishonest* (reverse coded), *distrustful of the agent/system–trustful of the agent/system*, *inconsiderate–considerate*, *confidential–divulging* (reverse coded), *not deceitful–deceitful* (reverse coded), and *disrespectful–respectful* ($\alpha = .92$). Responses to these nine items were averaged. Higher scores indicated higher trustworthiness ratings.

In the image conditions, the participants reported no prior exposure to any of the images. The five dependent measures of social responses either had no correlation or positive correlation with one another. Except that competency and trustworthiness had a strong correlation at $r(166) = .72$, no other two measures had very strong correlations, $r$s = .07–.57. Competency and trustworthiness have been conceptualized as two related but distinct constructs of credibility. Thus, high correlation is reasonable, and they remained as separate measures.

As an exploratory measure of the frame of comparison and the notion of default image for the text-output interface, the participants in the text condition answered a question: "If you had mentally pictured an image representing the system in the interaction, what would it have been?" The participants selected one out of the 12 facial images used in the other conditions as their answer. The 12 images were randomly positioned in a grid underneath this question. In addition, the participants answered *yes* or *no* to a second question: "Did you consciously picture any image that represented the system while you were interacting with the system?"

## 3. Results

Linear polynomial trend analyses were conducted to test the hypothesized linear relationship between the degree of anthropomorphism of computer representations and the five social response measures. The three images within each of the four image conditions were not differentiated because they were used to instantiate the corresponding anthropomorphism level. The text condition was not included in the linear trend analysis because it was not part of the anthropomorphism variation. Table 1 presents the means and standard deviations of the dependent measures in all the conditions.

The linear trend analysis within one-way ANOVA in SPSS v. 13, which uses contrast coefficients of −3, −1, 1, and 3 for the four anthropomorphism conditions, revealed significant linear trends for almost all the social response measures: for social judgment, $F(1,140) = 6.69$, $p < .01$, $\omega^2 = .03$; for homophily, $F(1,140) = 3.43$, $p = .07$, $\omega^2 = .01$; for social influence, $F(1,140) = 7.07$, $p < .01$, $\omega^2 = .03$; for competency, $F(1,140) = 10.18$, $p < .01$, $\omega^2 = .05$; for trustworthiness, $F(1,140) = 9.72$, $p < .01$, $\omega^2 = .04$. Accordingly, the means of the dependent measures showed linear increases from low to medium to high anthropomorphism levels and to real human images, except that the competency rating dropped slightly from the medium- to the high-anthropomorphism level. But this difference was not significant even through a liberal $t$-test, $t(70) = .48$, $p = .64$.

Table 1
Means and standard deviations of the social response measures in all conditions

| Measures | Low Mean (SD) | Medium Mean (SD) | High Mean (SD) | Human Mean (SD) | Text Mean (SD) |
|---|---|---|---|---|---|
| Social judgment | 4.02 | 4.15 | 4.28 | 4.85 | 4.63 |
| | (1.35) | (1.34) | (1.52) | (1.23) | (1.25) |
| Homophily | 3.16 | 3.35 | 3.55 | 3.79 | 3.95 |
| | (1.76) | (1.53) | (1.47) | (1.30) | (1.39) |
| Social influence | −.33 | −.30 | −.14 | −.11 | −.13 |
| | (.37) | (.36) | (.44) | (.47) | (.37) |
| Competency | 4.44 | 4.92 | 4.82 | 5.22 | 5.25 |
| | (1.06) | (.91) | (.87) | (.95) | (.93) |
| Trustworthiness | 4.25 | 4.73 | 4.77 | 5.02 | 4.69 |
| | (1.13) | (.86) | (.99) | (1.00) | (1.08) |

*Note:* Low, medium, and high refer to the levels of anthropomorphism of the computer-synthesized facial images. Human refers to the real human images. $N = 36$ in each of these four conditions. Text refers to the text-output condition which was not part of the linear trend analysis. $N = 24$ in the text condition.

Hence, as the facial images of the computer representations increased in their degree of anthropomorphism, they progressively received more social responses from people. People attributed them with more positive social judgment, greater homophily (approaching significance), and higher competency and trustworthiness ratings. In addition, higher anthropomorphism led to greater social influence. Thus, all five hypotheses received support. The results suggest an overall linear relationship between the degree of anthropomorphism of visual computer representations and people's social responses.

To explore how the text interface compared to the facial image interfaces, a Dunnett's test compared the text condition to the image conditions. The only significant comparison was that the text interface received higher competency rating than the low-anthropomorphism image interfaces, $p < .01$. Through examining mean similarity, the text condition seemed to resemble the human image condition or fell between the human image condition and the high-anthropomorphism condition.

The exploratory measure of deliberately asking the participants to pick a facial image to represent the text interface revealed that human images or high-anthropomorphism images were more likely to be picked than low- or medium-anthropomorphism images, approaching statistical significance, $\chi^2(3) = 6.33$, $p = .09$. Specifically, Images 10 and 7 (see Fig. 1) were chosen the most, respectively by 6 and 5 of the 24 participants in the text condition. However, 21 out of the 24 participants reported they did not consciously picture an image for the text interface during the interaction, $\chi^2(1) = 13.5$, $p < .001$. Therefore, the exploratory results about the text interface shed some light on the notion of default image, suggesting that the default might be a human image or a highly anthropomorphic image but such image attribution is most likely to be unconscious if it operates.

## 4. Discussion

The results of the study provided support to the often assumed linear relationship between the degree of anthropomorphism of computer representations and people's social responses. When facial representations on computers progressed from low-anthropomorphism to medium-anthropomorphism to high-anthropomorphism and to real human images, people gave them more positive social judgment, greater homophily attribution, higher competency and trustworthiness ratings, and were more influenced by them in choice dilemma decision-making. The collection of 610 computer-synthesized character images provided a large and externally valid pool to enable a three-level breakdown of the anthropomorphism degree. Multiple instantiations per level of anthropomorphism also helped provide a powerful test of anthropomorphism. All computer representations being still facial images solved the confounding of modality, which was common in previous studies. The control of physical attractiveness and facial expression in all nine synthetic images and the three real human images further boosted the internal validity of the results.

The notion of the linear relationship between the degree of anthropomorphism and social responses seems intuitive. The fundamental idea appears self-explanatory because people are both the ideal template of anthropomorphism and the seat of social attribution and responses. That is, the more a computer resembles *me* (i.e., human), the more I will respond in *my* (i.e., social) way. This basic notion, however, had resisted simple empirical verification. Among many reasons including the ones delineated earlier in this paper, multidimensionality in facial image processing and of anthropomorphism may also be responsible and beg more discussion.

Encountering a human image or a human-like image, many factors may operate. Prominent social identities such as race, gender and age, facial cues such as attractiveness, expressions and friendliness, and visual features such as style, dimensionality and aesthetics, may all play a role. To focus on the level of anthropomorphism, this research either controlled some of these factors or used multiple instantiations to dilute possible effects of other factors. The multiplicity of factors in facial image processing may be particularly an issue for a between-participants design, which was the case in almost all the studies in the literature including the present one. Although future research with a between-participants design can manipulate and contrast the level of anthropomorphism and other factors, a within-participants design may be valuable and more parsimonious by prominently and deliberately presenting variations of the degree of anthropomorphism.

The multiplicity of factors in facial image processing and the between-participants design may also explain the apparently small differences among the means of the dependent measures in the conditions of this study. This study aimed to prove the overall linear relationship based on the full range of anthropomorphism. Thus, comparison between any two levels was not called for. Future research concerning differences between any particular two levels of anthropomorphism will need to provide further theorization.

Anthropomorphism, which was not a salient concept in traditional studies of social sciences, is becoming increasingly important because computer-generated virtual social entities are propagating in the digital communication world. Being human-like, however, might potentially involve an array of dimensions defining humanness. Thinking, appearance and hearing, as basic human faculties, defined the efforts of technological anthropomorphism in artificial intelligence, visual realism and facial animation, speech recognition and synthesis. FtF human behavior has also naturally become the golden template for comparison. FtF communication is inherently multimodal. The multimodality of FtF communication may explain why prior efforts on testing the level of anthropomorphism included the modality factor. Future research will need to independently vary modality along with the level of anthropomorphism and pay attention to consistency between modalities.

Consistent with previous studies (Burgoon et al., 2000; Kiesler et al., 1996; Sproull et al., 1996), this study yielded relatively quite positive judgment of the text interface. A real human image or a highly anthropomorphic synthetic human image received the most pick when the participants were asked to select an image to retroactively represent the computer. The notion of default image might be better labeled as default persona, however, because most participants did not report to consciously picture an image for the text interface. This default persona is probably human. But this idea of human attribution is at best plausible and implicit because the participants' selection of a representative image was forced and confined by the provided images. Sundar and Nass (2000) have shown that computers are treated as an independent social actor, not as a representation of a programmer or another person over the network. Thus, it is unlikely that the default persona represents any real human being. It awaits more research with different methodological approaches to gain more understanding of implicit mental model of computers.

The text interface, nonetheless, was not perceived more positively than human images in this study or previous studies. But what might be discouraging to anthropomorphism technologists is that text interface appears superior to synthetic anthropomorphic visual representations in this and other studies. The issue is not whether images are better than text but what images are better than text, because human images are not less positive, if not more positive, than text. The hyperpersonal notion of idealized perception would not apply to text-interface in one-time human–computer interaction because hyperpersonal perception is claimed to be most likely to operate when relationship-building is existent or anticipated (Walther, 1996; Walther, Slovacek, & Tidwell, 2001). In computer-mediated interpersonal communication, Walther et al. (2001) found that presentation of the partner's photo boosted the intimacy/affection feeling and social attractiveness judgment in short-term interaction and only hurt such feeling and judgment in long-term interaction, compared to text-based interaction without photos. This and previous studies on anthropomorphism all incorporated one-time interaction. Hence, the "blame" is likely to come down to the particular synthetic facial images used.

Judgment of synthetic facial images hinges intrinsically on people's expectation of them. Expectation has received competing conceptualizations (Dehn & van Mulken, 2000; Nowak, 2004). One argument is when users see a synthetic human image representation on the computer, they expect other human-related capabilities from it. When these capabilities are not met, disappointment ensues. But this argument appears untenable because real human images in this study had identical behaviors as the synthetic ones.

Another argument is the notion of "uncanny valley" coined by roboticist Masahiro Mori (1982). It states that when a robot or a computer-synthesized anthropomorphic entity is extremely human-like subtle imperfections would stand out and drastically dampen people's judgment of it. In my judgment, however, none of the facial images used in this or other studies has reached that level. The "uncanny valley" thesis has been contested since the 1970s and has not received a precise conceptualization or empirical test.

A third argument is quite the opposite. The synthetic facial images used in this and other studies might not be good enough to parallel real human images. Sophistically generated synthetic characters such as those in high-budget computer animation films like "Shrek" are certainly impressive and do not appear to suffer from an uncanny valley effect. Thus, future research should test higher-end exemplars of anthropomorphism when technology for everyday computing further advances.

Also as anthropomorphic computer entities propagate, we should expect more repeated or long-term interaction between users and them. Very little research has examined such long-term interactions. How the degree of anthropomorphism influences long-term interactions will be even more complex. Users' expectation of the anthropomorphic entities should be higher given task or relational interdependency in long-term interactions. As a consequence, the effects of anthropomorphism on social responses as discerned in this study may be accentuated due to higher expectation and higher interests at stake. However, how long-term interactions evolve, specifically how expectations and fulfillments modify each other over time, should substantially moderate the effects of anthropomorphism. It awaits future systematic research to unpack the role of long-term exposure and interactions.

In conclusion, this research provided initial support to the notion that the more anthropomorphic a computer representation is, the more social people's responses are. It demonstrated variability proposed by Nass and Moon (2000): social responses can vary as a function of the degree of anthropomorphism of the computer representation. Social responses are certainly contextualized in specific interactions with computers. Understanding the basic social mechanism of responding to the degree of anthropomorphism, however, is essential for examining its operation in various human–computer interaction contexts such as entertainment, education, task performance, and collaboration. As computer-synthesized virtual social entities are becoming prevalent, a more refined conceptual framework of anthropomorphism is needed. In addition to visual images, future research should systematically incorporate other dimensions of computer representations such as voice, motion and other behaviors, but with careful conceptualization and execution of pairings between the dimensions. Then, we will have better understanding of how people perceive and respond to entities which are somewhat like us but not quite.

## References

Berlo, D. K., Lemert, J. B., & Mertz, R. J. (1970). Dimensions for evaluating the acceptability of message sources. *Public Opinion Quarterly, 33*, 563–576.

Burgoon, J. K., Bonito, J. A., Bengtsson, B., Cederberg, C., Lundeberg, M., & Allspach, L. (2000). Interactivity in human–computer interaction: A study of credibility, understanding, and influence. *Computers in Human Behavior, 16*, 553–574.

Cassell, J., Sullivan, J., Prevost, S., & Churchill, E. (2000). *Embodied conversational agents*. Cambridge, MA: MIT Press.

Dehn, D. M., & van Mulken, S. (2000). The impact of animated interface agents: A review of empirical research. *International Journal of Human–Computer Studies, 52*, 1–22.

Goetz, J., Kiesler, S., & Powers, A. (2003). Matching robot appearance and behavior to tasks to improve human–robot cooperation. In *Proceedings of the 12th IEEE international workshop on robot and human interactive communication (Ro-man 2003)*. Vol. IXX, Millbrae, CA. pp. 55–60.

Gong, L., & Lai, J. (2003). To mix or not to mix synthetic speech and human speech? Contrasting impact on judge-rated task performance versus self-rated performance and attitudinal responses. *International Journal of Speech Technology, 6*, 123–131.

Gong, L., & Nass, C. (2007). When a talking-face computer agent is half-human and half-humanoid: Human identity and consistency preference. *Human Communication Research, 33*, 163–193.

Kiesler, S., Sproull, L., & Waters, K. (1996). A prisoner's dilemma experiment on cooperation with people and human-like computers. *Journal of Personality and Social Psychology, 70*(1), 47–65.

Lee, E.-J. (2004). Effects of visual representation on social influence in computer-mediated communication: Experimental tests of the social identity model of deindividuation effects. *Human Communication Research, 30*, 234–259.

Lee, E.-J., & Nass, C. (2002). Experimental tests of normative group influence and representation effects in computer-mediated communication: When interacting via computers differs from interacting with computers. *Human Communication Research, 28*, 349–381.

Maes, P. (1994). Agents that reduce work and information overload. *Communication of the ACM, 37*(7), 31–40, p. 146.

McCroskey, J. C., Hamilton, P. R., & Weiner, A. M. (1974). The effect of interaction behavior on source credibility, homophily, and interpersonal attraction. *Human Communication Research, 1*, 42–52.

McCroskey, J. C., Richmond, V. P., & Daly, J. A. (1975). The development of a measure of perceived homophily in interpersonal communication. *Human Communication Research, 1*, 323–332.

Mori, M. (1982). *The buddha in the robot*. Tokyo, Japan: Charles E. Tuttle Co.

Nass, C., & Brave, S. (2005). *Wired for speech: How voice activates and advances the human–computer relationship*. Cambridge, MA: The MIT Press.

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues, 56*(1), 81–103.

Nowak, K. L. (2004). The influence of anthropomorphism and agency on social judgment in virtual environments. *Journal of Computer-Mediated Communication*, 9(2), Retrieved September 7, 2005 from http://jcmc.indiana.edu/vol2009/issue2002/nowak.html.

Nowak, K. L., & Biocca, F. (2003). The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. *Presence -Teleoperators and Virtual Environments, 12*(5), 481–494.

Ohanian, R. (1990). Construction and validation of a scale to measure celebrity endorsers' perceived expertise, trustworthiness, and attractiveness. *Journal of Advertising, 19*(3), 39–52.

Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers television and new media like real people and places*. New York, NY: Cambridge University Press.

Rice, R. E., & Love, G. (1987). Electronic emotion: Socioemotional content in a computer-mediated communication network. *Communication Research, 14*, 85–108.

Short, J., Williams, E., & Christie, B. (1976). *The social psychology of telecommunications*. New York: Wiley.

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

Sundar, S. S., & Nass, C. (2000). Source orientation in human–computer interaction: Programmer, networker, or independent social actor? *Communication Research, 27*, 683–703.

Turing, A. (1950). Computing machinery and intelligence. *Mind, 59*(236), 433–460.

Walther, J. B. (1996). Computer-mediated communication: Impersonal, interpersonal, and hyperpersonal interaction. *Communication Research, 23*, 3–43.

Walther, J. B., Slovacek, C. L., & Tidwell, L. C. (2001). Is a picture worth a thousand words? Photographic images in long-term and short-term computer-mediated communication. *Communication Research, 28*, 105–134.

Wheeless, L. R., & Grotz, J. (1977). The measurement of trust and its relationship to self-disclosure. *Human Communication Research, 3*, 250–257.