
Does it matter if a computer jokes?

Peter Khooshabeh

University of Southern California
Institute for Creative Technologies
12015 E. Waterfront Dr.
Playa Vista, CA 90094 USA
khooshabeh@ict.usc.edu

Cade McCall

Max Planck Institute for Human
Cognition and Brain Sciences
Stephanstraße 1A, 04103
Leipzig, Germany
mccall@cbs.mpg.de

Sudeep Gandhe

University of Southern California
Institute for Creative Technologies
12015 E. Waterfront Dr.
Playa Vista, CA 90094 USA
gandhe@ict.usc.edu

Jonathan Gratch

University of Southern California
Institute for Creative Technologies
12015 E. Waterfront Dr.
Playa Vista, CA 90094 USA
gratch@ict.usc.edu

James Blascovich

UC Santa Barbara
Department of Psychology
Santa Barbara, CA 93106-9660
blascovi@psych.ucsb.edu

Abstract

"We need oxygen, especially if someone farts!"

The goal here was to determine whether computer interfaces are capable of social influence via humor. Users interacted with a natural language capable virtual agent that told persuasive information, and they were given the option to use information from the dialogue in order to complete a problem-solving task. Individuals interacting with an ostensibly humorous virtual agent were influenced by it such that those who judged the agent unfunny were less likely to be persuaded and departed from the agent's suggestions. We discuss the implications of these results for HCI involving natural language systems and virtual agents.

Keywords

humor, joy of use, virtual agents, natural language processing, social influence, persuasion, ontology

ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous; H1. Models and Principles; K.4 Computers and Society

General Terms

Design, Human Factors, Management, Performance, Theory

Copyright is held by the author/owner(s).

CHI 2011, May 7–12, 2011, Vancouver, BC, Canada.

ACM 978-1-4503-0268-5/11/05.

Introduction

Individuals often try to influence others. Whether in business or casual contexts, people use various techniques to exert social influence. For example, sales people have the daunting task of motivating consumers to buy an assortment of products that they often do not need. And, consumers fall prey to effective social influence strategies. In more casual settings, an individual might want to influence her friends to accept her suggestion of where to eat dinner or what film to see.

There are many successful social influence strategies but an important one involves interpersonal attraction or “liking.” Cialdini suggests that when people like a potential influencer, they are more likely to follow her suggestions [2]. Besides physical appearances that lead to liking, there are also social and behavioral ones including humor.

One function of humor is to ‘break the ice’ in a social situation. Hence, it is not surprising that people employ humor across different settings, including persuasive ones, even though the humorous dialogue does not necessarily attempt to communicate a persuasive message. In many cases, people try to be humorous simply to put other individuals in a positive affective state. For example, in a highly stressful situation, jokes contextualized around the situation might alleviate distress and allow people to persevere on situational tasks. Individuals can use jokes in order to build positive affect and trust in others so as to influence and ultimately persuade them to implement a specific course of action like buying a car or signing a petition or doing things in a certain way.

In this paper, we review work from the behavioral sciences on the effects of humor on cognition and decision-making. We frame our discussion with respect to HCI applications and discuss our study of humor in a decision making task involving a virtual agent. The research question is whether a humorous virtual agent can influence or persuade users to make decisions in the Lunar Survival Scenario task [6]. In brief, the Lunar Survival Scenario informs the participant that she is stranded on the moon and has to decide how to prioritize items that will help save her life (see Table 1). This task: 1) presents a stressful situation in which humor can alleviate some of the pressure and 2) the task allows us to quantify how much social influence the virtual agent has on the participant.

The results suggest that users who judged a virtual agent to be funny also tended to be influenced by that agent. Moreover, if the virtual agent made an above average number of humorous assertions, then it was more effective at influencing users. This contribution is novel because the study used a natural language capable virtual agent. The spontaneous nature of humor makes interactive natural language generation essential to really capitalize on mirth. Given the current state of the art, it is difficult to get perfect natural language understanding, but, nonetheless, the interactive natural language capable agent produced context appropriate humorous language often enough to have social effects.

Related Work

Although there is a fair amount of literature on the study of humor [9], there is less work using humor as a social influence strategy; a survey of the article titles from the Journal of Humor from 1988-2009 shows that

only one looked at persuasion [1]. Business scholars report that humor in the work place facilitates cooperation and affinity in corporate teams [3]. Negotiation is commonly practiced in business and various other domains and an important aspect of it is to influence others to reach an agreement.

In the context of real-world dyadic negotiations, social psychologists have conducted controlled experiments using human confederates who made attempts at humor. O'Quin and Aronoff [12] had participants play the role of an art purchaser. Experimental confederates playing the role of the seller made a task irrelevant humorous comment with respect to selling a painting, such as, "My final offer is \$1000, *and I'll throw in my pet frog.*" Confederates in the non-humorous condition just made the monetary offer. People in the humorous condition were more likely to accept the seller's offer.

Ideal Rank Objects	Order for Sample Participant Rankings	Difference Score
Oxygen Tank (1)	Oxygen (2)	1
FM Receiver (5)	Receiver (8)	3
First Aid Kit (7)	First Aid (6)	1
Parachute (8)	Parachute (9)	1
Flares (10)	Flares (10)	0
Pistols (11)	Pistols (15)	4
Compass (14)	Compass (3)	11

table 1. The first column shows some of the objects in the Lunar Survival Scenario and their order based on rankings from NASA. Sample rankings made by participants are in the second column, followed by the difference in rankings.

Relatively few HCI researchers have studied how humorous dialogue with a digital agent affects task performance and perceptions of the technology (e.g., computer) itself. Morkes, Kernal, and Nass [11] had participants rank order items in the Desert Survival Scenario. In this task, participants are asked to prioritize items, such as a flashlight and a plastic raincoat, in order of importance to survive on a desert. In Study 1, Morkes et al. told the participants that they were interacting with another person participating in the task over the network. In fact, they were actually chatting with a computer that had preprogrammed responses about the Desert Survival Scenario items. For half the participants, the preprogrammed responses contained non-offensive jokes that were related to the task but not relevant. Humor participants reported greater participation and liked the "other person" more than those in the non-humor condition. Study 2 was identical to the first but the participants were told that they were chatting with a computer. Interestingly, although the preprogrammed responses were identical, a comparison across the two experiments showed that the HCI humor participants in Study 2 were less sociable and spent less time on the task.

It is important to point out that the human-computer dialogue in the work by Morkes and colleagues was based on preprogrammed comments. Regardless of what each participant typed, the computer made the same comments in the same order for every participant. Therefore, one of the contributions of our work is that we are using interactive natural language processing instead of preprogrammed responses. We were interested to see whether humor would exert social influence when the virtual agent is not simply

responding to the user based on a preprogrammed dialogue but rather a more natural, dynamic one.

Aside from humor, other HCI researchers have used a similar ranking task to study whether tailoring a messenger of suggestive information can have specific social effects, such as influence. In the health domain, Yin, Bickmore, and Cortes [14] designed embodied conversational agents that resembled users' ethnic identities by making one appear Latino and another Anglo-American. They also varied whether the agents spoke Spanish or English (implemented via a run-time text-to-speech engine). Their virtual agent made assertions about the pros and cons of exercise. Users had the task of ranking assertions about exercise before and after chatting with the agent. Yin et al. found that individual user characteristics determined whether the culturally congruent agent influenced users' attitudes toward exercise.

Based on the literature review, we hypothesized that a humorous virtual agent will exert greater social influence than a non-humorous virtual agent. The theoretical mechanism underlying our hypothesis is that users will like the funny virtual agent more, and will more readily accept its suggestions in the Lunar Survival Scenario task because they like the funny agent [10].

We conducted the experiment using an interactive natural language system tailored for the Lunar Survival Scenario using the Domain Editor [9, 10]. This tool allowed us to start building a natural language system from the top down. The first step was specifying the items involved in the Lunar Survival Scenario. The natural language knowledge base for the virtual agent

represented the rankings suggested by NASA experts for the Lunar Survival Scenario. The next step was to assign surface text utterances to the complete set of speech acts that the Domain Editor generated. Finally, we conducted iterative user testing to improve the natural language understanding.

We hypothesized that the humorous virtual agent will lead users to make rankings similar to it, leading to a higher "social influence" score, which we defined as the difference between the pre-chat and post-chat user rankings. For example, users receive a high social influence score if they rank the items very differently from the NASA rank order prior to chatting with the virtual agent and then rank more similar to the NASA rankings suggested by the virtual agent.

Method

Participants

Undergraduate psychology students (N=54) voluntarily consented to be a part of the experiment in order to fulfill course credit. There were 12 males and 42 females who were on average 18 years of age ($SD = .7$). One participant was excluded from analysis because the post-chat rankings score (76) was larger than two SD's away from the mean.

Design

The study used a between subjects design with humor as the independent variable ($n = 27$ for each condition). Dependent measures were defined as performance on the object rankings in the Lunar Survival Scenario and responses to a post-survey.

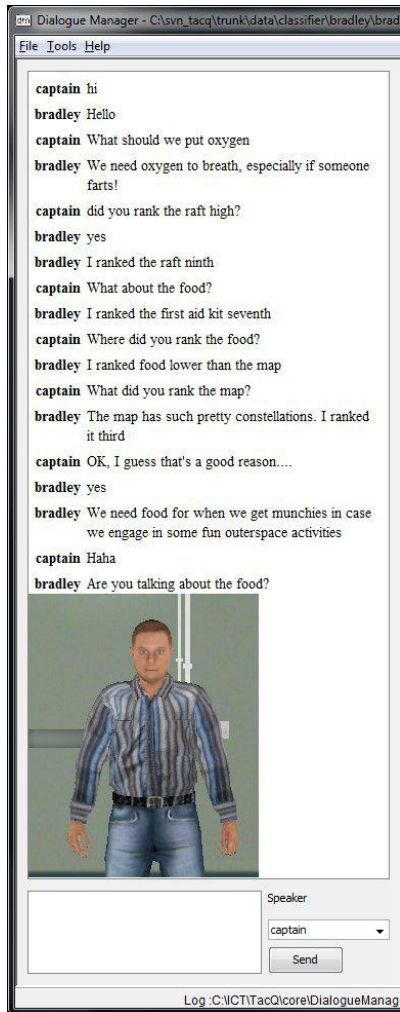


figure 1. Sample interactive chat

Apparatus

We created the virtual agent's knowledge base for the Lunar Survival Scenario [6] using the Domain Editor tool [9, 10]. We specified all the relevant objects in the task (see Table 1), and assigned them values with respect to the rank order suggested by NASA experts. The Domain Editor generated all the relevant speech acts and our task was to assign plausible utterances to the speech acts. Through this process we designed the virtual agent to suggest the NASA rank order for the items in the Lunar Survival Scenario.

Materials

The basic demographic survey contained questions about computer experience as well as personality scales, which included the positive and negative affect schedule (PANAS) [13] and items adapted from the social co-presence, interactant satisfaction, and emotional intelligence scales [8].

The Lunar Survival Scenario [6] presented participants with 15 items they had to rank order. Performance was measured by how much participants' rankings for each item deviated from the rankings suggested by NASA experts; this absolute deviation was summed to represent one difference score (see Table 1).

Using the Domain Editor, we assigned humorous assertions to some of the dialogue acts for the humorous agent. The agent's suggestions about the rank order for the items are always based on those suggested by NASA experts. An example of a humorous assertion regarding a question about the FM receiver is, "We can use the FM receiver to communicate with another ship, or we can pass time with some fun music

on the radio." Figure 1 shows an example interactive session with the humorous virtual agent.

Procedure

During the Lunar Survival Scenario experimental task, participants were informed that they have crash landed on the moon and need to choose items in order to trek 200 km back to a life-saving rendezvous point. Additionally, they are told that another crew member is also present, whose name is Bradley. Due to the crash landing, the captain has been incapacitated and the participant is now the officer in charge. Bradley, the other crew member, knows the ship's inventory well. Participants are told that Bradley is a virtual agent with whom they can chat, and that he is a non-native English speaker. We did this so that participants would tolerate the agent when the natural language understanding inevitably failed and the virtual agent had to ask the participants to clarify their question.

Although the other crew member knows the inventory well, the instructions and the experimenter stress that the participant is the officer in charge. As the captain, the participant had to make the final decision on how to rank the items. We intended the participants to have the decision power and responsibility in the scenario so that they would not blindly rely on the advice from the virtual agent.

Participants initially ranked items prior to chatting with the virtual agent by dragging and dropping the task items on a graphical user interface. After the participants completed their pre-chat rankings, the experimenter reminded the participants that they were in charge and now had an opportunity to chat with Bradley. Participants had to decide how they would use

the information that Bradley told them in order to make a set of post-chat rankings.

Results

Manipulation check

As Figure 2 shows, not all individuals in the humorous condition thought the agent was funny. Similarly, about 30% of the participants in the non-humorous condition thought the agent was funny. However, three independent raters performed verbal protocol analyses of the conversation logs between participants and the virtual agent (*Cronbach's alpha* = .87). These coders rated the conversation logs with the humorous version of Bradley as more humorous ($M = 5.7$, $SD = 1.8$ vs. $M = 3.5$, $SD = 1.5$) based on a holistic code for the entire conversation, which was a 7-point likert scale question, $t(51) = 4.9$, $p < .001$.

Effect of chatting with virtual agent

We quantified the Lunar Survival Scenario rankings via the deviation of the participants' scores from the rankings suggested by NASA experts. Perfect rankings according to this scheme would result in a zero difference, and the worst possible ranking would result in a difference score of 210. Participants in each condition scored similarly on their pre-chat rankings, $t(51) < 1$, $p = .42$.

A 2×2 repeated measures ANOVA with humor condition (funny or not) and ranking task (pre vs. post-chat rankings) as independent variables revealed a main effect for ranking task, $F(1, 49) = 40.3$, $p < .001$, $\eta^2 = .45$. Participants post-chat rankings ($M = 35$, $SD = 12$) became more similar to the agent's rankings compared to their rankings prior to chatting ($M = 46$, $SD = 9.5$), see Table 2. However, no effect of

humor condition, $F < 1$, nor an interaction between humor condition and ranking task were significant, $F(1, 49) = 1.11$, $p = .3$.

	Humorous	Non-humorous
Pre-chat rankings	45.4 (9.9)	47.1 (9.1)
Post-chat rankings	36.5 (12.5)	33.4 (12.2)

table 2: The main effect showing that individuals performed better on the task as a result of chatting with a virtual agent.

Does humorous natural language influence participant responses?

One of the contributions of our work is a system that is capable of interactive humorous natural language as opposed to merely a pre-crafted dialogue with scripted humorous comments [e.g., 6]. A reason why the humorous agent may not have had more influence was because not all humorous agents had a chance to deliver their humorous utterances. Based on a priori predictions, we analyzed the number of unique humorous assertions that the virtual agent made and how it affected its social influence on the users. We computed the social influence score by subtracting the post-chat score from the pre-chat score. This difference would be a large positive number if post-chat ranks are lower, which indicates that participants ranked more similarly to the virtual agent after chatting with it.

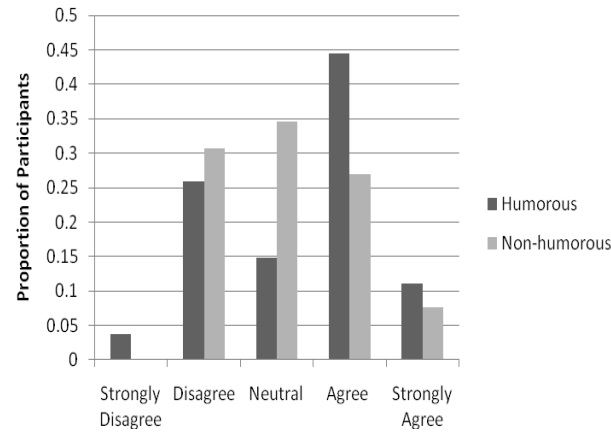


figure 2. Histogram of the proportion of participants in each condition and how they responded to the likert scale item that asked whether they judged the agent to be funny

For example, if someone adopts the same exact rankings as the virtual agent, then their post-chat ranking difference would be zero. Suppose that this individual's pre-chat ranking was 46, which is the mean of the pre-chat ranking scores. If we then subtract the post-chat ranking score from the pre-chat ranking score, the social influence score is 46. Therefore, larger positive numbers indicate that participants made post-chat rankings more similar to the agent.

Unique humorous assertions were normally distributed ranging from 1 to 9 of them ($M = 4.4$, $SD = 2.1$). We performed a median split on the number of unique humorous assertions. An independent samples t test showed that if the virtual agent made more than 4 unique humorous assertions, it was more effective at

influencing the users rankings, $t(18) = 1.98$, $p < .05$ (see Figure 3).

The histogram distribution in Figure 2 shows that not all participants in the humorous condition thought that the agent was funny. Consequently, we examined the correlation between the subjective evaluation of humor and the social influence score.

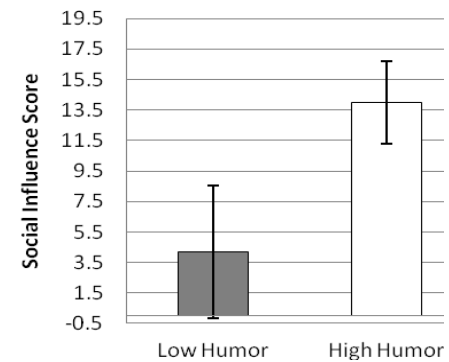


figure 3: The effect of unique humorous assertions on social influence.

We predicted that the social influence scores should be correlated with the extent to which participants judged the virtual agent as humorous. This would suggest that individuals who judged the virtual agent as humorous would also tend to be influenced more. For the participants in the humorous condition, there was a correlation between the social influence scores and how funny they judged the agent, $r(26) = .56$, $p = .003$ (see Figure 4). For the participants in the non-humorous condition, there was no correlation between social influence scores and how funny they judged the agent, $r(25) = .24$, $p = .24$.

The correlational analysis suggests that individuals who perceived the virtual agent as being intentionally humorous were more influenced to adopt the agent's ranking suggestions. To further test this inference, we performed a median split on the participants' ratings of how funny they judged the virtual agent. This eliminated all the individuals who were neutral about whether they judged the virtual agent as humorous (4 in the humorous and 9 in the non-humorous condition). An independent samples t test on the social influence scores for those in the humorous condition showed that participants who judged the agent as humorous ($M = 14.7$, $SD = 8.4$) were more influenced than those who did not think that the virtual agent was humorous ($M = -2.3$, $SD = 6.6$), $t(20) = 4.9$, $p < .001$. The same independent samples t test for the non-humorous condition showed that there was no difference based on whether they judged the agent as humorous or not, $t(15) < 1$, $p = .43$.

Subjective Evaluations of the Humorous Agent

In this paper we focused the analysis on the social co-presence scales; we aggregated subscales of the co-presence, interactant satisfaction, and emotional credibility scales ($\alpha = .72$). A multivariate ANOVA revealed that humor had a significant effect on the subjective evaluations, *Wilk's Lambda* (5, 41) = 5.1, $p < .001$. Follow-up univariate ANOVA's on the subscales revealed that users of the humorous agent evaluated it as more friendly, $F(1, 45) = 4.5$, $p < .05$, $\eta^2 = .09$, as using feelings, $F(1, 45) = 7.1$, $p < .05$, $\eta^2 = .14$, as showing interest, $F(1, 45) = 21.9$, $p < .001$, $\eta^2 = .33$, and being more intimate, $F(1, 45) = 7.7$, $p < .01$, $\eta^2 = .15$, compared to users of the non-humorous agent. Conversely, users of the non-humorous agent felt that

the agent was more detached, $F(1, 45) = 6.5$, $p < .05$, $\eta^2 = .13$ (see Figure 5).

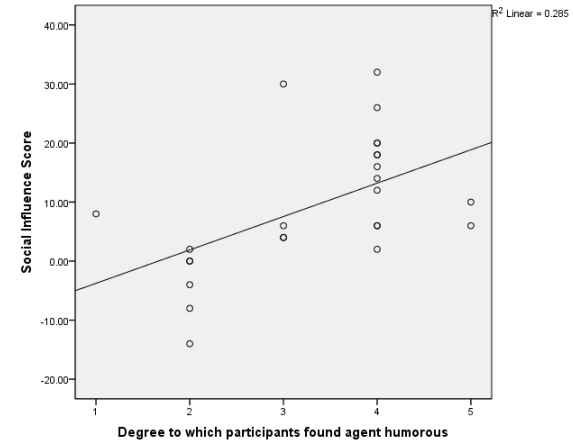


figure 4: Scatter plot showing correlation between social influence and humor perception in the funny condition.

Discussion

In a controlled experiment, we demonstrated that a humorous virtual agent, when judged as funny, is more effective at socially influencing users. Moreover, the effect is driven by the number of unique humorous assertions that arise during the natural language interaction. When the virtual agent made an above average number of unique humorous assertions, it influenced users more effectively.

One surprising result is that a sizable proportion of users of the non-humorous virtual agent judged it as funny. Conversely, some users in the humorous virtual agent condition did not think it was funny. We can explain the latter result as being attributed to the subjective nature of humor. There is some evidence

from that suggests that humor is less effective on individuals whose need for cognition is high [15]. It might be possible to personalize humor interfaces to individuals whose need for cognition is low and to avoid using humor interfaces with individuals whose need for cognition is high.

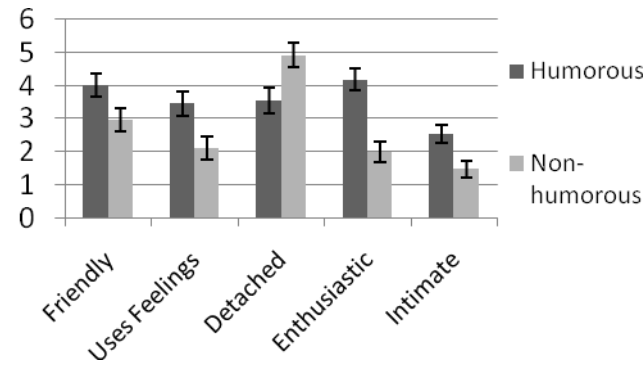


figure 5. Subjective evaluations of the different virtual agents.

It is unclear why users of the non-humorous agent would find it funny and this is a question for future research. More importantly, we can speculate that humor inoculated the virtual agent against user aggression and dissatisfaction and possibly led users to perceive the virtual agent as witty and clever. When the users interacted with the virtual agent and it generated more unique humorous assertions, then they were more likely to adopt its suggestions. Also, the subjective evaluations of the virtual agent overall, regardless of the number of unique humorous assertions the agent generated, suggests that users were generally more positive towards the funny agent and actually perceived the non-humorous agent as detached from the task.

With respect to the previous work by Morkes and colleagues, our results suggest that the interactive humorous natural language is important for social influence purposes. Whereas Morkes et al used a scripted dialogue, our virtual agent was interactive and responded to questions that users asked.

Our results lend support to the *liking* theory of social influence. Users were more attracted to the humorous virtual agent and they identified with it more. The humorous virtual agent was more successful at socially influencing users who liked it. However, it is also theoretically plausible that the humor led users to redefine the situation as less threatening and therefore more accepting of the humorous virtual agent's suggestions. We hope to investigate these theoretical alternatives in future work.

The results have potential implications for user interface design strategies. For example, designers of recommender systems can explore humorous verbal communication for social influence purposes. Herlocker and colleagues [7] review the current state of the art in collaborative filtering algorithms for recommender systems and focus on how to evaluate such systems. They particularly emphasize keeping users' goals in mind and choosing the correct task, evaluation metrics, and appropriate datasets. Given that at some level the goal of a recommender system is to influence a user's choice, recommendations contextualized with humorous assertions could influence how users respond.

Participants saw only a static picture of the virtual agent (see Figure 1). In future work, we are interested in how an animated character will affect social influence and the subjective evaluations. Finally, we are

interested in studying whether it matters when virtual agents make humorous assertions using realistic voices. Our future studies will incorporate naturalistic text-to-speech in order to address the question of whether the modality of humor affects social influence.

Acknowledgements

We thank the reviewers for their suggestions and criticisms, the summer 2010 ICT interns for helping to pilot test the dialogue system, and the UCSB research assistants for help conducting the experiment. Peter Khooshabeh performed this research while on appointment as an Oak Ridge Associated Universities Postdoctoral Fellow with the Army Research Laboratory, Human Research and Engineering Directorate.

References

1. Articles published in Humor, 1988-2009. *International Society for Humor Studies, Journal of Humor Research*, 2011.
http://www.hnu.edu/ishs/ISHS/Documents/Humor1988_2009.pdf.
2. Cialdini, R.B. *Influence: the psychology of persuasion*. Harper Collins, New York, NY, 2007.
3. Clouse, R.W. and Spurgeon, K.L. Corporate analysis of humor. *Psychology: A journal of human behavior* 32, (1995), 1-24.
4. Gandhe, S., DeVault, D., Roque, A., et al. From domain specification to virtual humans: an integrated approach to authoring tactical questioning characters. *Proceedings of Interspeech*, (2008).
5. Gandhe, S., Whitman, N., Traum, D., and Artstein, R. An integrated authoring tool for tactical questioning dialogue systems. *Proceedings of 6th Workshop on Knowledge and Reasoning in Practical Dialogue Systems*, (2009).
6. Hall, J. *NASA Moon Survival Task: The original consensus exercise*. Teleometrics International, The Woodlands, TX, 1989.
7. Herlocker, J.L., Konstan, J.A., Terveen, L.G., and Riedl, J.T. Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems (TOIS)* 22, 1 (2004), 5-53.
8. Kang, S., Watt, J.H., and Ala, S.K. Social copresence in anonymous social interaction using a mobile video telephone. *Proceeding of the Twenty-Sixth Annual SIGCHI Conference on Human Factors in Computing Systems*, ACM (2008), 1535-1544.
9. Martin, R.A. *The psychology of humor: an integrative approach*. Elsevier Academic Press, London, 2007.
10. Mettee, D.R., Hrelec, E.S., and Wilkens, P.C. Humor as an interpersonal asset and liability. *The Journal of Social Psychology* 85, 1, 51-64.
11. Morkes, J., Kernal, H.K., and Nass, C. Effects of humor in task-oriented human-computer interaction and computer-mediated communication: a direct test of SRCT theory. *Human-Computer Interaction* 14, 4 (1999), 395-435.
12. O'Quinn, K. and Aronoff, J. Humor as a technique of social influence. *Social Psychology Quarterly* 44, 4 (1981), 349-357.
13. Watson, D., Tellegen, A., and Clark, L.A. Development and validation of brief measures of positive and negative affects: The PANAS scale. *Journal of Personality and Social Psychology* 54, (1988), 1063-1070.
14. Yin, L., T. W. Bickmore, and Cortes, D.E. The impact of linguistic and cultural congruity on persuasion by conversational agents. *Proceedings of IVA, Lecture Notes in Computer Science*, Springer (2010), 343-349.
15. Zhang, Y. Responses to humorous advertising: the moderating effect of need for cognition. *The Journal of Advertising* 25, 1 (1996), 15-32.