Proceedings of the Thirtieth International Florida Artificial Intelligence Research Society Conference

Listen to My Body: Does Making Friends Help Influence People?

Ron Artstein, David Traum,
Jill Boberg, Alesia Gainer, Jonathan Gratch,
Emmanuel Johnson, Anton Leuski
USC Institute for Creative Technologies
12015 Waterfront Dr., Playa Vista CA 90094 USA

Mikio Nakano
Honda Research Institute Japan Co., Ltd.
8-1 Honcho, Wako, Saitama 351-0188 Japan

Abstract

We investigate the effect of relational dialogue on creating rapport and exerting social influence in human-robot conversation, by comparing interactions with and without a relational component, and with different agent types. Human participants interact with two agents – a Nao robot and a virtual human – in four dialogue scenarios: one involving building familiarity, and three involving sharing information and persuasion in item-ranking tasks. Results show that both agents influence human decision-making; people prefer interacting with the robot, feel higher rapport with the robot, and believe the robot has more influence; and that objective influence of the agent on the person is increased by building familiarity, but is not significantly different between the agents.

Introduction

Conversational robots working in offices and at home are expected to establish *rapport* with human users through repeated interaction and personalization (Kanda et al. 2012; Lee et al. 2012). The present study measures the effects of relational dialogue on both subjective measures of rapport and objective measures of social influence, in order to establish design principles for conversational robots that can build rapport with users.

In dialogue system research, many have contrasted *functional* or *task-oriented* dialogue, aimed at joint completion of a specific task, with *relational* or *social* dialogue, aimed at building a relationship between the participants. There are different types of functional dialogue: some functions relate to information-seeking or service tasks, but in others, a goal of the system is to persuade or influence the user, rather than just divine and comply with their intent (Traum et al. 2005; Manuvinakurike, Bickmore, and Velicer 2014; Hiraoka et al. 2014). For these sorts of tasks it would seem that relational dialogue might play an instrumental role in the persuasive function.

The objective of our research is to examine the importance of combining relational and functional dialogue for conversational robots, for a wide variety of objective and subjective evaluation criteria. In this paper we describe the first step: an experiment that examines the impact of relational dialogue on both subjective and objective measures, namely, the participants' perceived rapport and the influence exerted in persuasive dialogue tasks. Our experiment considers two factors: the presence or absence of relational dialogue, and the type of agent, whether a robot or a virtual human (2×2 design). We find an effect of agent type on perceived rapport, and of relational dialogue on social influence. In the next section, we review social influence and factors that are correlated with it. This is followed by a description of our agents, the specific goals and design of our experiment, the experiment protocol and dependent measures, results of the experiment, and a discussion.

Related Work

Social Influence

Social influence arises across many contexts; here we study it using the classical 'survival task,' where members of a team rank the importance of a number of items (e.g., items that might help one survive a crash in the desert). These are ranked individually and then re-ranked following a team discussion, and the difference in ranking provides an objective measure of social influence. Besides providing an objective measure of persuasion, the wide adoption and extensive empirical analysis of this task allows one to compare findings across different theoretical perspectives and different fields of research. Survival tasks have been used in psychology (Wang et al. 2015), communication studies (Takayama, Groom, and Nass 2009), virtual human research (Khooshabeh et al. 2011), and human-robot interaction (Adalgeirsson and Breazeal 2010).

Familiarity

One factor that shapes social influence in general, and survival tasks in particular, is the familiarity team members have with each other. In general, people are less influenced by strangers or people they feel more distant from. A standard way to overcome this obstacle to persuasion is to 'break the ice' between teammates. For example, Behrend, Whelan, and Thompson (2008) showed that a short ice-breaker (consisting of simple questions such as 'where do you live?') significantly increased rapport and social influence within the team. We are unaware of similar attempts to use ice-breakers with machine teammates, however Khooshabeh et al. (2011) showed that telling jokes could increase persua-

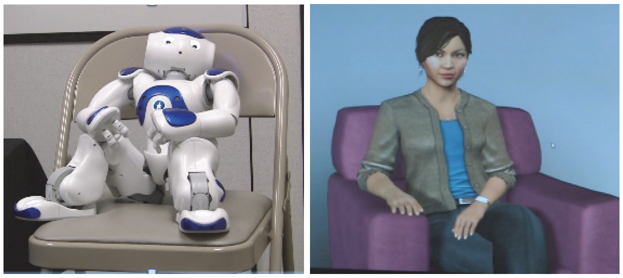

Figure 1: Artificial Agents Niki and Julie

sion in a lunar survival task, and several studies suggest that various rapport-building techniques can enhance human-machine teamwork (Bailenson and Yee 2005; Gratch et al. 2007). We use relational dialogue to establish familiarity.

Embodiment

Another factor that impacts social influence in general, and the survival task in particular, is embodiment. Psychological and communication studies suggest that embodiment increases social influence. For example, participants were less persuaded in a lunar survival task when their teammates communicated via teleconference compared with face-to-face interaction (Werkhoven, Schraagen, and Punte 2001).

Findings in human-machine interaction are more mixed. Many studies show increases in subjective engagement but not objective persuasion. For example, Adalgeirsson and Breazeal (2010) compared a fully-articulated robot to a static interface in the desert survival task and found increased subjective presence, trust, cooperation and engagement, but no objective persuasion. Powers et al. (2007) compared a collocated robot, a remote robot projected on a large screen, and a computer agent projected on a monitor or a large screen. Robots were more engaging than agents; that is, people spent more time with robots, but they disclosed their personal information more to the agents, and least to the collocated robot. There was no significant difference in influence on behaviors. Similarly, Takayama, Groom, and Nass (2009) compared a robot with a disembodied voice in the desert survival task and found people liked the robot significantly more when it agreed with them, liked it significantly less when it disagreed, but found no objective difference in influence. Some robotic studies have found objective benefits to embodiment (Kiesler et al. 2008; Leyzberg et al. 2012), but not in the context of a survival task.

Agents

We use two agents in the experiment: a robot named Niki and a virtual human named Julie (Figure 1). Niki is a Nao, a humanoid robot commonly used in human robot interaction studies (Deshmukh et al. 2015; Dominey et al. 2010); Julie is based on the Virtual Human Toolkit (Hartholt et al. 2013). Julie is presented in two modes: when engaging in relational dialogue she is presented multimodally with voice and virtual embodiment on a computer monitor, including non-verbal behaviors such as lip movements, facial expressions and head and hand gestures; in the task dialogues she is presented in voice only, as if through a teleconference. Motivation for the latter is because unimodal spoken dialogue is a common way of communicating with agents, necessitated by some use cases such as while driving, cooking or operating machinery.

The agents differ on several dimensions besides physical embodiment: Niki is presented as male and appears child-like, while Julie is designed as an adult female; Niki is clearly a man-made artifact while Julie looks almost photo-real; they have different voices, and their gesturing capabilities are different. Differences in the measured outcomes could be attributed to any or all of these underlying differences between the agents. Thus, our experiment does not attempt to isolate an abstract notion of embodiment while keeping all other aspects constant; rather, we compare two popular and accessible types of agents, each with state of the art abilities for its type. While many studies in the human-robot interaction community have looked at physical vs. virtual robots, we think a more relevant choice is between commonly used platforms that may differ in a wide variety of features, even if it is not completely clear which set of features make a difference.

The agents share a common set of hand-authored text utterances, designed specifically for this experiment; from these utterances, individual behaviors with synchronized gestures and speech were generated for each platform. Both systems use a collection of beat gestures and nodding as they communicate; a total of 5 gestures are used for the robot, and each gesture is authored in a way to mimic the expressivity of the virtual human. Niki came standard with an on board text-to-speech engine and thus it was used to synthesize the robot's speech. All of Niki's behaviors were authored using Choregraphe, a multi platform application which allows users to create complex behavior for the Nao robot, and each behavior was hand synced to align the speech with gesture. Julie's voice was synthesized using a voice from NEOspeech's text-to-speech engine. Her behaviors were generated with synchronized speech and gesture using Smartbody, the Nonverbal Behavior Generator, and TTSRelay, all components of the Virtual Human Toolkit. The behaviors of both agents were constructed off-line, and stored for use during the experiment.

Both the robot and virtual human are operated in a "Wizard of Oz" fashion, where a human operator issues commands to execute agent behaviors using a push-button GUI. The interface communicates with the agents using the VHMsg[1] system from the Virtual Human Toolkit, sending messages which tell the agents which of the stored behaviors to execute in real time.

Experimental Design

As mentioned in the introduction, our goal is to examine the impact of relational dialogue on both perceived rapport and influence in persuasive dialogue tasks. We broke this down into the following concrete research questions.

[1] https://sourceforge.net/projects/vhmsg/

Q1. Can artificial agents influence human decision-making through dialogue?

Q2. Does the embodied nature of the agents (e.g. robot vs. disembodied voice) matter for influence?

Q3. Can rapport with artificial agents be increased through face-to-face dialogue interaction?

Q4. Does building familiarity impact influence?

Q5. Does the level of rapport correlate with influence?

Q6. How stable is rapport across influence tasks?

Our expectation from previous literature is that the answers to Q1, Q3 and Q4 will be yes, while it is less clear about Q2, Q5, and Q6.

The remainder of this section describes the design of the experiment; subsequent sections present the specific procedure and results.

Tasks

Participants engaged in four dialogues with one or both of the agents. Three dialogues were ranking tasks designed to measure social influence (two classical "survival tasks" and an original "Save the Art" task based on a similar model), and one dialogue was an ice-breaker designed to create familiarity. The dialogues were always conducted in the following order:

1. One of the survival tasks with one of the agents.
2. The ice-breaker dialogue with the second agent.
3. The other survival task, again with the second agent.
4. The Save the Art task with both agents.

The survival tasks asked participants to rank the importance of specific items in terms of their usefulness for survival after landing in a hostile environment (in our case, the moon and the desert). After initially ranking the items, the participant and the agent discussed their rankings, and then the participant re-ranked the items. The agents' rankings were pre-determined, and were the same for each of the agents.

The Save the Art task asked the participants to rank pieces of art in the order in which the pieces should be saved from a fire in an art museum; this was followed by a discussion with both agents and re-ranking the items. The task differed from the survival tasks in some ways. First, the ranking of art pieces is subjective, as opposed to the relative objectivity of the survival task, in which the agents' rankings were based on expert consensus about impact on survival. Additionally, the discussion was between the participant and both agents, each of whom had their own unique (and fixed) ranking of the items (i.e., three-party persuasion dialogue). These rankings were structured to reflect a similar distance from an "average" participant ranking.

Between the first and second survival tasks was an ice-breaker dialogue, which involved a semi-structured conversation between the agent and the participant. After exchanging greetings, the agent began asking the participant a series of seven open-ended questions, such as, "Where are you from?", "What is your favorite kind of music or favorite music artist?", and "Have you traveled?". The agent commented on the participant's answers, and revealed short anecdotes about themselves on the same subjects.

Non-familiar	Familiar	Familiar	Both agents
Robot, Desert	Ice-breaker	Voice, Lunar	Save the Art
Robot, Lunar	Ice-breaker	Voice, Desert	Save the Art
Voice, Desert	Ice-breaker	Robot, Lunar	Save the Art
Voice, Lunar	Ice-breaker	Robot, Desert	Save the Art

Table 1: The four experimental conditions

The ice-breaker was the only dialogue in which the virtual human was animated; in the ranking tasks, the virtual human was displayed as a static image on a the screen. Therefore, participants whose ice-breaker dialogue was conducted with the robot never saw the virtual human animated, and only knew her as a voice.

Variables

The experiment has two primary independent variables: *agent* and *familiarity*. Agent is just the agent itself, either a robot or a voice. Familiarity is created through the ice-breaker dialogue: the familiar agent is the one who has previously engaged in the ice-breaker with the participant. Since each participant interacted with both the robot and the voice agent, with one of the two being familiar and the other non-familiar, we are able to treat either agent or familiarity as a repeated measure.

The *task* was treated as a random variable: the desert survival task and lunar survival task could be assigned to either agent, the only restriction being that each participant engaged in a different survival task with each agent. This resulted in a total of 4 experimental conditions (Table 1). We did not control for the *ordering* of the tasks separately from familiarity: the participant's first survival task was always with the non-familiar agent, followed by the ice-breaker dialogue and a survival task with the familiar agent. Since the participant's rankings were taken before and after each task, an additional variable of *time* applies to the rankings (before/after the interaction).

There are three dependent variables: *divergence* and *influence* are objective measures of the difference in item rankings between participant and agent, and changes in these rankings before and after interaction; *perceived rapport* is a subjective measure obtained from the participant through a questionnaire. These are described in detail below.

Method

Participants and Setup

There were 40 participants in the study, recruited through Craigslist[2] (16 male, 24 female; ages 24–70, median 44). Participants were paid $40 for their effort. Each participant was randomly assigned to one of four experimental conditions (Table 1); there were 10 participants in each condition.

[2] http://craigslist.org

	Niki created a sense of closeness or camaraderie between us.
R	Niki created a sense of distance between us.
	I think that Niki and I understood each other.
R	Niki communicated coldness rather than warmth.
	Niki was warm and caring.
R	I wanted to maintain a sense of distance between us.
	I felt I had a connection with Niki.
	Niki was respectful to me.
R	I felt I had no connection with Niki.
	I tried to create a sense of closeness or camaraderie between us.
R	I tried to communicate coldness rather than warmth.

Figure 2: Rapport questionnaire. Each question is rated on a 5-point scale. Items were always presented in the same order. The label **R** indicates reverse-coded items.

Please answer the following questions about each of the characters you interacted with:

- How trustworthy were the characters?
- How persuasive were the characters?
- How natural did you find the characters' voices?
- How natural did you find the characters' movements?

Which character would you prefer to interact with again in the future?

What did you like about Niki?

What did you dislike about Niki?

What did you like about Julie?

What did you dislike about Julie?

Figure 3: General questions about the interaction

Participants were seated at a table, holding an iPad Pro, facing a robot to their right and a screen with speakers to their left. To avoid distraction, the robot was covered when not in use. Participants wore a close-talking microphone which recorded their speech throughout the interaction, and a video camera placed between the robot and the screen recorded their face and upper body movements. Starting with the 32nd participant, a second video camera was added behind the participant to capture the entire scene.

Procedure

After filling the consent forms, participants were given the iPad Pro running a survey using a platform provided by Qualtrics,[3] which guided them through the experiment. Participants advanced through the survey screens, and their responses were recorded by the survey software.

For each of the ranking tasks, participants started by reading the instructions, and then ranked the items on the iPad Pro. After they finished the ranking, they had a conversation with the agent (Robot, Voice, or both), during which they could rank the items again (the participant's previous rankings were not visible). After the interaction, participants had the opportunity to revise their ranking before submitting it, and then filled out the 11-item rapport questionnaire (Figure 2; von der Pütten et al. 2010). For the Save the Art task participants filled out the rapport questionnaire twice – once for each agent. For the ice-breaker task there were no items to rate; participants had a conversation with the agent, and then filled out the rapport questionnaire. After the final task, participants filled out some general questions (Figure 3) and demographic information on the iPad Pro; they were then debriefed and paid.

Measures Taken

The primary measurements are the item rankings and the rapport questionnaire, measured by the survey software. The rapport questionnaire is straightforward – the responses are averaged to yield a single number which reflects the participant's perception of their rapport with the agent after the task. The item rankings are used to infer the amount of influence the interaction with the agent had on the participant. For each participant ranking recorded, we calculate *divergence* as the Kendall τ distance between the participant's ranking and the agent's ranking (Euclidean distance yields roughly the same results). We infer *influence* by comparing divergences before and after the interaction: if the rankings are closer (divergence is lower) after the interaction, then we know that the influence was positive – the agent convinced the participant to change their rankings in the direction of the agent's ranking. Note that divergence and influence are symmetric – they could just as well be used to measure the influence of the participant on the agent. However, since by design the agent doesn't change its rankings, we know that this measures the influence of the agent on the participant.

Unfortunately, it is difficult to compare divergence and influence across tasks, because of systematic differences in the initial divergence: participants' initial rankings in the Desert survival task were substantially further away from the agents' rankings than in the Lunar survival or Save the Art tasks, which means there was much more room to converge in the Desert survival task. We therefore introduce measures of *normalized divergence* and *normalized influence*: the effect of task is neutralized by dividing a measure (influence or divergence) by the mean of all users and agents on the same task, and then multiplied by the mean on all tasks in order to bring it back to the original scale (scaling back is necessary in order to retain the difference between divergence measures before and after the interaction).

Results

Influence

The agents were able to influence the participants, across all tasks: an ANOVA measuring divergence as a function of task, agent, familiarity, and time (before/after the interaction) showed a highly significant main effect of time ($F(1,296) = 219, p < 10^{-15}$): divergence was consistently higher before each interaction than after it. This drop in divergence (positive influence) held for each of the 3 tasks separately (Desert survival: $t(39) = 12, p < 10^{-13}$; Lunar survival: $t(39) = 8.8, p < 10^{-10}$; Save the Art: $t(39) = 5.8, p < 10^{-6}$). The ANOVA also showed a highly significant main

[3] https://www.qualtrics.com

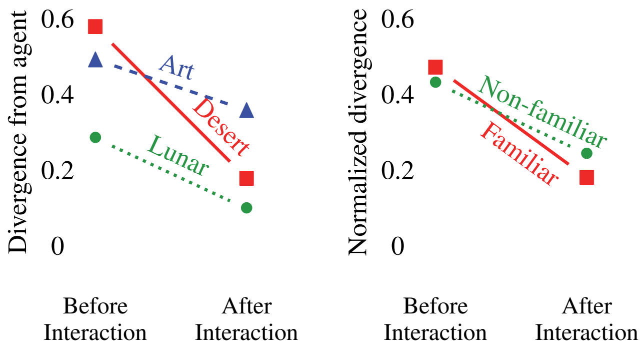

Figure 4: (left) Participant rankings are closer to the agent rankings after the interaction, indicating that the agent is able to influence the participants. (right) In the survival tasks, participants are influenced more by the familiar agent.

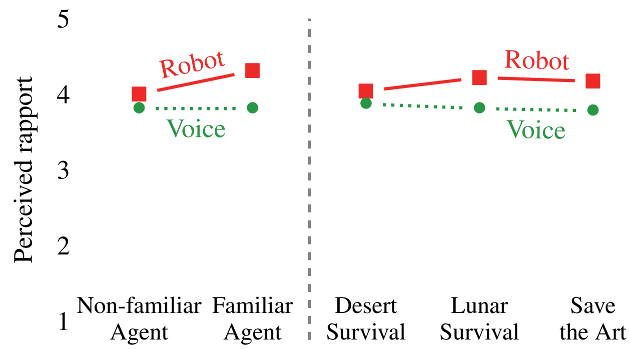

Figure 5: (left) People feel more rapport with the robot than with the voice; people also feel more rapport with the robot when the robot is the familiar agent. (right) Perceived rapport does not appear to be influenced by the task.

effect of task ($F(2,296) = 89, p < 10^{-15}$) and an interaction between time and task ($F(2,296) = 29, p < 10^{-11}$), indicating that both divergence and influence systematically differed between tasks (Figure 4, left): *yes* to Q1. No other effects or interactions were significant: *no* to Q2.

While task was intended as a random variable, the large differences in divergence and influence between the tasks make it difficult to detect the effects of the other experimental conditions. To neutralize the effect of task we used *normalized influence*; this can also be used as a repeated measure in the survival tasks, since each participant had two survival interactions – one with the non-familiar agent and one with the familiar one. Taken as a repeated measure, we find a significant effect of familiarity on normalized influence, with agents in the familiar condition showing higher influence on the participants ($t(39) = 2.9, p < 0.006$, two-tailed; Figure 4, right): *yes* to Q4. As noted above, there is a confound between familiarity and order, since participants always interacted with the familiar agent after interacting with the non-familiar agent. We did not find a significant effect of agent on influence, nor did we find an effect of familiarity on influence in the Save the Art task.

Rapport

Unlike influence, perceived rapport differed by agent: an ANOVA measuring perceived rapport as a function of agent and familiarity showed a significant main effect of agent ($F(1,76) = 6.7, p < 0.05$). This effect was much stronger as a repeated measure ($t(39) = 3.8, p < 0.001$, two-tailed): participants reported higher rapport with the robot than with the virtual human. There was no main effect of familiarity, and the interaction between agent and familiarity was non-significant. There was a trend of familiarity in the robot condition alone: participants reported higher rapport with the robot when the robot was the familiar agent ($t(37) = 1.91, p = 0.06$, two-tailed; Figure 5, left): *unclear* to Q3. A separate ANOVA revealed a significant interaction between familiarity and task ($F(2,148) = 3.7, p < 0.05$): participants reported higher rapport with the familiar agent in the Lunar task only. There was no significant interaction between agent and task, suggesting that perceived rapport is fairly stable across the influence tasks (Figure 5, right): *yes* to Q6.

Interestingly, we did not find any correlation between influence and perceived rapport. Familiarity affects influence, and the ice-breaker task was initially intended as a rapport-building measure. However, familiarity does not affect perceived rapport, and the two dependent measures are not correlated: *no* to Q5.

Other Measures

The other subjective measures in Figure 3 indicate that more participants prefer the physical embodiment with Niki to the virtual human Julie. Three quarters of the participants prefer to interact again with Niki, also Niki was trusted more and thought to be more persuasive. Julie's voice and movements had higher averages on naturalness, though this was not significant.

In the open-ended questions, participants characterized Niki as responsive, engaging and with a personality, noted that talking to him was like talking to a child, commented on his physical presence and movement, and also noted his intelligent responses. What participants mostly didn't like about Niki was his voice, which some participants found hard to understand. Julie was liked primarily for her intelligence, pleasantness and warmth, and for her voice; participants did not like her lack of animation (in the ranking tasks), noted that she was less engaging than Niki, and commented on her distance, and the feeling that she was less natural.

Conclusions and Future Work

This paper reported the results of our investigation into the impact of relational dialogue on the perceived rapport and exerted influence in decision-making tasks, and how the impact changes depending on the embodiment by comparing a Nao robot and a disembodied voice agent that is embodied as a virtual human only when engaging in relational dialogue. The experimental setting is novel in employing a

2×2 design that separates embodiment and familiarity as distinct independent variables. This enabled us to investigate how those factors influence human decision-making and perceived rapport, although these have not been considerably examined in previous work. We have obtained several interesting findings including (a) the participants felt they had higher rapport with the robot than with the virtual human, (b) relational dialogues influence participants' decision-making, (c) the types of embodiment (that we used in the experiment) do not impact influence, and (d) the level of perceived rapport does not impact influence. These suggest that the type of embodiment is important in building rapport but that rapport does not always impact influence on human decision-making.

Although we think these suggestions are useful in designing tasks and interactions of conversational robots and agents, there remain many open questions. As a next step, we plan to analyze the dialogue collected in this study in more detail to investigate the effects of the agents' micro-behaviors on human decision-making in the discussions in the re-ranking tasks. We also plan to conduct a cross-cultural study because the participants in this study are all native American-English speakers living in the US and the results may not be the same for other people. Finally we will investigate the influence of rapport in types of decision-making tasks other than re-ranking.

Acknowledgments

This study was funded by Honda Research Institute Japan Co., Ltd.

The first seven authors were supported in part by the U.S. Army; statements and opinions expressed do not necessarily reflect the position or the policy of the United States Government, and no official endorsement should be inferred.

References

Adalgeirsson, S. O., and Breazeal, C. 2010. Mebot: A robotic platform for socially embodied presence. In *Proc. HRI*, 15–22. IEEE Press.

Bailenson, J., and Yee, N. 2005. Digital chameleons: Automatic assimilation of nonverbal gestures in immersive virtual environments. *Psychological Science* 16:814–819.

Behrend, T. S.; Whelan, T. J.; and Thompson, L. F. 2008. Kickoff meetings for computer-mediated teams: Effects on team performance. In *23rd annual conference of the Society for Industrial and Organizational Psychology*.

Deshmukh, A.; Jones, A.; Janarthanam, S.; Hastie, H.; Ribeiro, T.; Aylett, R.; Paiva, A.; and Castellano, G. 2015. An empathic robotic tutor in a map application. In *Proc. AAMAS 2015*, 1923–1924.

Dominey, P. F.; van der Zant, T.; Lallee, S.; Jouen, A.-L.; Hinaut, X.; Weitzenfeld, A.; van Hoof, H.; and Dávila-Chacón, J. 2010. Cooperative human robot interaction with the Nao humanoid: Technical description paper for the "radical dudes". In *RoboCup@Home Technical Description Papers: Germany - Singapore*.

Gratch, J.; Wang, N.; Gerten, J.; Fast, E.; and Duffy, R. 2007. Creating rapport with virtual agents. In *Proc. IVA 2007*. Springer. 125–138.

Hartholt, A.; Traum, D.; Marsella, S. C.; Shapiro, A.; Stratou, G.; Leuski, A.; Morency, L.-P.; and Gratch, J. 2013. All together now: Introducing the Virtual Human Toolkit. In *Proc. IVA 2013*. Springer. 368–381.

Hiraoka, T.; Neubig, G.; Sakti, S.; Toda, T.; and Nakamura, S. 2014. Reinforcement learning of cooperative persuasive dialogue policies using framing. In *COLING*, 1706–1717.

Kanda, T.; Shiomi, M.; Miyashita, Z.; Ishiguro, H.; and Hagita, N. 2012. An affective guide robot in a shopping mall. In *Proc. HRI*, 173–180. IEEE.

Khooshabeh, P.; McCall, C.; Gandhe, S.; Gratch, J.; and Blascovich, J. 2011. Does it matter if a computer jokes? In *CHI'11 Extended Abstracts on Human Factors in Computing Systems*, 77–86. ACM.

Kiesler, S.; Powers, A.; Fussell, S. R.; and Torrey, C. 2008. Anthropomorphic interactions with a robot and robot-like agent. *Social Cognition* 26(2):169.

Lee, M. K.; Forlizzi, J.; Kiesler, S.; Rybski, P.; Antanitis, J.; and Savetsila, S. 2012. Personalization in HRI: A longitudinal field experiment. In *Proc. HRI*, 319–326. IEEE.

Leyzberg, D.; Spaulding, S.; Toneva, M.; and Scassellati, B. 2012. The physical presence of a robot tutor increases cognitive learning gains. In *Proc. COGSCI*, 1882–1887.

Manuvinakurike, R.; Bickmore, T. W.; and Velicer, W. 2014. Automated indexing of internet stories for health behavior change: Weight loss attitude pilot study. *Journal of Medical Internet Research*.

Powers, A.; Kiesler, S.; Fussell, S.; and Torrey, C. 2007. Comparing a computer agent with a humanoid robot. In *Proc. HRI*, 145–152. IEEE.

Takayama, L.; Groom, V.; and Nass, C. 2009. I'm sorry, Dave: I'm afraid I won't do that: Social aspects of human-agent conflict. In *Proc. SIGCHI*, 2099–2108. ACM.

Traum, D.; Swartout, W.; Marsella, S.; and Gratch, J. 2005. Fight, flight, or negotiate: Believable strategies for conversing under crisis. In *Proc. IVA 2005*. Springer. 52–64.

von der Pütten, A. M.; Krämer, N. C.; Gratch, J.; and Kang, S.-H. 2010. "It doesn't matter what you are!" Explaining social effects of agents and avatars. *Computers in Human Behavior* 26(6):1641–1650.

Wang, Y.; Lucas, G.; Khooshabeh, P.; de Melo, C.; and Gratch, J. 2015. Effects of emotional expressions on persuasion. *Social Influence* 10(4):236–249.

Werkhoven, P. J.; Schraagen, J. M.; and Punte, P. A. J. 2001. Seeing is believing: Communication performance under isotropic teleconferencing conditions. *Displays* 22(4):137–149.