Culture, Errors, and Rapport-building Dialogue in Social Agents

Gale M. Lucas, Jill Boberg, David Traum, Ron Artstein, Jonathan Gratch, Alesia Gainer, Emmanuel Johnson, Anton Leuski
USC Institute for Creative Technologies
Los Angeles, CA
United States
lucas@ict.usc.edu

Mikio Nakano
Honda Research Institute Japan Co., Ltd.
Wako, Saitama
Japan
nakano@jp.honda-ri.com

ABSTRACT

This work explores whether culture impacts the extent to which social dialogue can mitigate (or exacerbate) the loss of trust caused when agents make conversational errors. Our study uses an agent designed to persuade users to agree with its rankings on two tasks. Participants from the U.S. and Japan completed our study. We perform two manipulations: (1) The presence of conversational errors – the agent exhibited errors in the second task or not; (2) The presence of social dialogue – between the two tasks, users either engaged in a social dialogue with the agent or completed a control task. Replicating previous research, conversational errors reduce the agent's influence. However, we found that culture matters: there was a marginally significant three-way interaction with culture, presence of social dialogue, and presence of errors. The pattern of results suggests that, for American participants, social dialogue backfired if it is followed by errors, presumably because it extends the period of good performance, creating a stronger contrast effect with the subsequent errors. However, for Japanese participants, social dialogue if anything mitigates the detrimental effect of errors; the negative effect of errors is only seen in the absence of a social dialogue. Agent design should therefore take the culture of the intended users into consideration when considering use of social dialogue to bolster agents against conversational errors.

CCS CONCEPTS

• **Human-centered computing → Human computer interaction (HCI)** → Empirical studies in HCI

KEYWORDS

Social agents, influence, social dialogue, rapport, conversational errors, culture

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
IVA '18, November 5–8, 2018, Thornleigh, NSW, Australia
© 2018 Association for Computing Machinery.
ACM ISBN 978-1-4503-6013-5/18/11…$15.00
https://doi.org/10.1145/3267851.3267887

ACM Reference format:

G. M. Lucas, J. Boberg, D. Traum, R. Artstein, J. Gratch, A. Gainer, E. Johnson, A. Leuski and M. Nakano. 2018. Culture, Errors, and Rapport-building Dialogue in Social Agents. In IVA '18: 2018 ACM International Conference on Intelligent Virtual Agents, November 5–8, 2018, Sydney, Australia. ACM, NY, NY, USA, 8 pages. https://doi.org/10.1145/3267851.3267887

1 INTRODUCTION

Human-agent interaction often involves engaging in multiple tasks, where each task has different expectations for the level of performance or communication errors. Designers of agents are faced with a question of ordering: is it better to put the more error-prone tasks first or last? And what other activities, for example rapport-building [1], can ameliorate (or exacerbate) the effects of errors? In previous work we examined the interplay between rapport and conversational errors, and found that rapport built through an ice-breaker, if it extends a period of good performance, can actually exacerbate the detrimental effect of errors that occur in a later interaction [2]. This result was found with American participants; the present paper extends the study to Japanese, and finds that their response pattern is different, suggesting that culture also plays a role. This investigation contributes to a long-term goal of understanding how rapport-building in conversations with agents can improve or worsen conversational outcomes, and under what circumstances these outcomes occur.

Previous work has shown that errors are particularly problematic for agents designed to have social influence over their users ("persuasive agents"). Indeed, when such agents make conversational errors, they are less capable of influencing people than agents that do not make errors [3,4]. Previous research also suggests that errors occurring after a period of good performance are more harmful to influence than those that occur earlier [5]. This was the motivation for our previous work, which tested whether building of rapport through social dialogue would mitigate or exacerbate the detrimental effect of an agent's errors on its ability to influence people, and how the effect of social dialogue interacts with the timing of the conversational errors [2]. Our present study considers the same two experimental factors (the presence of errors and the presence of an ice-breaker dialogue), together with a novel exploratory factor, the role of culture (thus, ultimately, a 2x2x2 design). Impact is measured using both objective measures of social influence and subjective measures of the participants' perceived rapport with the agent. We find that the ice-breaker dialogue exacerbates the effects of

errors among Americans, reducing how they find the agent and how much rapport they feel. However, we find that this "backfire effect" of rapport-building only occurs for Americans. Japanese participants, in contrast, are more persuaded if the agent uses an ice-breaker to build rapport. The findings have practical implications for the design of systems for human-computer interaction. The utility of most persuasive agents is, by definition, determined by their ability to build relationships and influence human users. The remainder of the paper presents background on influence, rapport and conversational errors; describes the experiment; and presents the results and conclusions.

2 RELATED WORK

2.1 Social dialogue

Dialogue system research often makes a distinction between *functional* or *task-oriented* dialogue, aimed at joint completion of a specific task, and *relational* or *social* dialogue, aimed at building a relationship between the participants. While some dialogue systems for intelligent virtual agents attempt to engage in only one of these types, others have included both relational and functional aspects. For example, both kinds of dialogue were used in the REA system [6], as well as the SASO system [7]. There are different types of functional dialogue – some functions relate to information-seeking or service tasks (assistant systems), but others focus on influencing the user. Examples include assistive systems to motivate people to exercise [8] or to negotiate on a course of action [7,9]. When influence is an objective, it is important that the user feels both comfortable getting input from the system and acting on that input. This trust is a critically important factor when attempting to create and maintain a relationship with an agent [10-11].

2.2 Persuasion and rapport

Some agents are able to persuade humans to change their behavior, and persuasive agents are explicitly designed to do so. For example, participants who received more social feedback from agents are more likely to conserve energy [12-13]. Virtual agents are also useful in persuading users to engage in behavior change around health [8, 14-18].

To effectively engage and persuade users, agents often try to build rapport with their users [19]. Rapport between two individuals creates interpersonal responsiveness and influence [20], so when rapport is created, research suggests that it can improve the quality of customer-employee exchanges [21], therapist-patient relationships [22], college roommate relationships [23], teacher-student interactions [24], relationships between caregivers and their charges [25], as well as other relationships. Similarly, rapport between agents and humans can build a relationship between the two, as well as help agents to influence human behavior. One study found that verbal cues of expertise, for example, increase the persuasiveness of an agent [26], while another found that human-agent teams that engaged in argumentation-based dialogue for shared decision-making were more successful than those where the agent engaged only in supervisory dialogue [27]. In other studies, participants who received social feedback were influenced to use less electricity in their homes or office than those with less social feedback [12-13].

One way to build rapport with users is through social dialogue. There are many different kinds of social relational techniques that use social dialogue to build rapport and trust, including as small talk, self-disclosure (intimacy), expert's jargon (credibility), gossip (social networks), and politeness (dignity/prestige) [6, p.91]. While these kinds of social dialogue have been used by humans to build trust with each other, they could also be used by agents when it is crucial that the agent gains the user's trust (e.g., persuasion tasks). Another kind of social dialogue that serves this purpose is an "ice-breaker." An icebreaker involves self-disclosure, but not necessarily intimate self-disclosure. In an ice-breaker, people share information about themselves with each other. Ice-breakers, specifically, are designed to increase trust and build rapport between group members, and often used in education and organizational fields for this purpose [28–30]. This is within the capabilities of agents – studies have shown that users are responsive to agents' verbal and non-verbal cues, social language, conversational gestures, and mutual gaze [6,31-32].

2.3 Conversational errors and trust

For the foreseeable future, social agents (including persuasive agents) will make occasional conversational errors. Designers who build such agents have to decide *whether* and *when* a new task that is expected to have a relatively high error rate should be given to the user. First, to the question of *whether* such a task should be given to users, work with virtual agents has shown that errors in dialogue systems can reduce social influence [3,4]. However, there is some mixed evidence. Within a human-agent team, trust is found to be most influenced by agents' characteristics such as reliability [10] and trust is easily lost but not easy to regain [5]. Other work finds that errors do not impact agents' influence – for example, one study found that errors affected the perceived reliability and trustworthiness of the agent, but had no impact on users' willingness to comply with the agent's requests [11]. It is possible, however, that the nature of this task lent itself to sustained compliance with the agent's request in spite of these errors, which occurred later in the task. In contrast, when other tasks are used (as in [33-35]), errors that occur after periods of good performance interfere with agents' influence. As such, in the present work, we use a standardized task to measure influence.

Second, to the question of *when* tasks likely to have high error rates should be included, studies have found that drops in reliability after a period of good performance were much more harmful to trust and performance than early failures, after which trust could be at least partially recovered [33-34]. For example, the reliability of an agent that shifted from 100% reliable to 80% reliable was rated lower than an agent that was 80% reliable from the start [5]. As such, we would expect that, when errors occur late in the dialogue, they would be detrimental to persuasion.

Recent research has also considered the impact of agents' social dialogue on the impact of errors. Our own research [2] has tested the possibility that building rapport through a social dialogue (e.g., ice-breaker) could mitigate the detrimental effect of errors on influence. However, ice-breaker *did not* benefit influence given that errors occur later; an error-free initial task plus an ice-breaker resulted in an extended period of good performance, which (as mentioned above) has been shown in other work to render later errors more detrimental [31].

Participants in [2] seemed to lose trust when the errors began after such a long period of good performance, and thus rapport built in the ice-breaker was not enough to mitigate the detrimental effect of errors. However, given this research was done with Americans, it is possible that –in different cultures– a *benefit* of social dialogue could instead be seen even when errors occur late. In more interdependent cultures (like East Asia), rapport-building is essential to negotiation; because rapport-building may be seen as prerequisite to a task in such cultures, engaging in a rapport-building ice-breaker might buffer against errors in such cultures.

3 EXPERIMENT

3.1 Design

To test for the effect of errors and social dialogue as in [2], we had each participant engage in two tasks with an embodied agent; errors were either present or absent in the second task. In between the tasks was either an ice-breaker dialogue or a control task. As described below, participants were recruited in either the U.S. or Japan. Thus, the experiment represented a 2 (error: errors or no errors) x 2 (ice-breaker: ice-breaker or control) x 2 (culture: U.S. or Japan) design. Participants were in one of these 8 conditions.

3.2 Participants

We recruited 209 participants (50.8% female). The U.S. sample was a subset of the sample reported in [2]. This sample was recruited from Craigslist. The Japanese sample was recruited by a professional recruiting service. They were compensated for their participation (30 USD or 7,273 JPY, respectively). Participants' age ranged from 18 to 84 with a mean of 31.4 years. U.S. participants were all native English speakers and all Japanese participants were native Japanese speakers. The data for 11 users had to be excluded due to technical issues experienced during the session. Analyses were conducted on the remaining 198 participants.

3.3 Procedure

Participants were seated at a table facing the humanoid physically embodied agent (NAO brand). The agent used a synthetic voice. Agent utterances were in English for Americans, and were translated into Japanese for the Japanese sample (see section 3.4). Participants wore a close-talking microphone which recorded their speech throughout the interaction, and a video camera recorded their face and upper body movements. A second video camera behind the participant recorded the embodied agent and its movements. To avoid distraction, the embodied agent was covered when not in use. The participants used a tablet for ranking items in the tasks, answering survey questions, and reading instructions on each phase of the experiment.

Participants started with the Lunar Survival Task, a problem-solving task widely used for measuring persuasion [2, 35–37]. In this task, participants were asked to imagine that they are part of a space crew that crashed on the moon, and were asked to rank 10 items as to their importance for surviving long enough to be rescued. Participants first ranked these items individually on the tablet, and were then told that they should rank them again with the help of another crewmember, our agent. At that time, the embodied agent was uncovered and activated, standing up from a crouched position, and participants engaged in dialogue with the agent. Although participants were led to believe that the agent was autonomous, in fact it was controlled by a human operator ("Wizard of Oz") and acted as a confederate [38], providing factual arguments for ranking the items in a specific order. The agent had a fixed set of arguments, all supporting a specific ranking order, but each argument was used as needed in the conversation with a given participant. Following the dialogue, participants re-ranked the items; the differences between initial rankings and final rankings served as a measure of influence [2, 39]. The embodied agent was again deactivated and covered, and the participant reported how much rapport they currently felt with the agent by filling a questionnaire. This scale was designed and used previously to measure the feeling of having a close and harmonious relationship, even after a single interaction [2, 40].

After the Lunar Survival task, participants answered a series of personal questions. In the ice-breaker condition, the questions were delivered through interactive dialogue with the agent. The agent shared its own stories, in first person, while eliciting stories from the participant. Questions asked of the participant included, "What is your favorite kind of music or favorite music artist," with follow up questions such as, "What other type of music do you like," or "Have you travelled?" followed by "Where was your favorite vacation?" and "Can you tell me more about your trip?". In the control condition, the user participated in a non-interactive oral survey, with the same personal questions being asked by a female (non-robotic) voice, with no feedback or other sharing from the voice. Audio for the control task came from a small set of speakers that were placed adjacent to the agent.

Participants then completed a second ranking task, the Save-the-Art Task [2, 39]. [1] Participants were asked to imagine that they were a manager at a museum that was on fire. They were asked to rank 10 pieces of art as to their importance in being saved. They ranked the items individually, had an interactive dialogue with the agent about which items should be saved, and then re-ranked the items. Although the Save-the-Art Task is not

"Can you repeat that? I didn't understand you"
Wrong answer
No answer
"I'm sorry, I didn't hear you"
Repeat and/or interrupt
Answer with a question or a non sequitur
"Can you repeat that? I didn't understand you" (× 3)
Wrong answer
No answer
"I'm sorry, I didn't hear you" (× 3)
Wrong Answer
Repeat/interrupt
Answer with question or non sequitur
"Can you repeat that? I didn't understand you" (× 2)
No answer

Figure 1: *Error manipulations, in order used.*

[1] The order of the tasks was not randomized, as it was shown previously that task order of these kinds of survival tasks did not affect results significantly [39].

	Japanese	English Translation	Error type
Robot:	どれを 2 番にしましたか？	What did you rank second?	
User:	ん，じゃーあーの東海道五十三次ー歌川広重の木版画です．	Well then it's the Fifty-Three Stations of the Tokaido, a woodcut print by Hiroshige Utagawa.	
Robot:	マンビラ木像．	Mambila Figure.	**Wrong answer**
User:	マンビラ木像はいちー，一番	Mambila Figure is the – the first.	
Robot:	はい．	Yes.	
Robot:	私は籠に入った花を 2 番にしました．その時代に女性が書いた絵は少ないので，価値があるからです．	I ranked Basket of Flowers second because there are few paintings by women from that time and that makes it worth.	
User:	うん．	Yeah.	
User:	じゃーそれはーその次にします．	Then I will rank it — next.	
Robot:	<long silence>	<long silence>	**No answer**
User:	籠に入った花は 3 番目にします．	I will rank Basket of Flowers third	
Robot:	はい．	Yes.	
Robot:	どれを 3 番にしましたか？	What did you rank third?	
User:	籠に入った花です．	I ranked Basket of Flowers third.	
Robot:	すみません．聞き取れませんでした．	I' m sorry. I didn't hear you.	**"I'm sorry, I didn't hear you"**

Figure 2: Example dialogue with errors.

a survival task – the rankings are more about objective and subjective analysis of the value of art pieces themselves – previous research had found that the pattern of influence of the two tasks was similar [2, 39]. As such, differences between rankings indexed influence. Participants again reported how much rapport they currently felt with the agent on the same scale [2, 40].

In the error condition, the agent made a series of errors while interacting with the participant on the Save-the-Art Task. Errors were introduced in one of several ways: asking users to repeat themselves, answering a different question than the user had asked, repeating the answer to the previous question, answering with a question or non-sequitur, or not answering at all. Errors were introduced into the dialogue according to a set order at a rate of about one of these errors per two utterances (Fig. 1). In most cases, the entire list was not used, as the task finished after about 10 errors. See Fig. 2 for an example of part of a dialog with errors.

3.4 Translation

All questionnaires and utterances were translated from the original English into Japanese for Japanese participants. To preserve meaning, Japanese translations were natural, rather than literal. The gestures made by the NAO while speaking were almost the same in both locations. In both cultures, there were minor glitches due to the synthesized voice: the intonation of some words were strange, and at times in the Japanese version there were shorter pauses at punctuation than would be expected.

4 RESULTS

In each of the ranking tasks (Lunar Survival and Save-the-Art), the item rankings were used to infer the amount of influence the interaction with the agent had on the participant. To do so, we calculated divergence between the participant's ranking and the agent's ranking, and compared the divergences before and after the interaction. If the rankings were closer after the interaction, we take that as a sign that the agent convinced the participant to change their rankings to be closer to its own (positive influence). Divergence is calculated as the Kendall τ distance between the participant's ranking and the agent's ranking, and ranges from 0 (identical rankings) to 45 (maximally different rankings); influence therefore ranges from −45 to 45, with larger numbers indicating more influence, and zero indicating no influence (negative numbers indicate that the participant moved farther away from the agents' ranking after the interaction).

4.1 Task 1

A 2 (error) x 2 (ice-breaker) x 2 (culture) ANCOVA was run on influence in task 1 (Lunar Survival Task). Because initial agreement with the agent limits the amount of possible influence on a given task, initial agreement with the agent on this task was entered as a co-variate to statistically equate participants on this factor. As expected, there was no effect of either manipulation, nor an interaction (Fs < 1.74, ps > .19). As the manipulations had not yet occurred, this showed that failure of random assignment did *not* occur (i.e., groups were equivalent). However, there was a significant effect of culture such that Japanese were overall less influenced than Americans ($F(1,189) = 13.33$, $p < .001$; see Fig. 3).

Likewise, a 2 (error) x 2 (ice-breaker) x 2 (culture) ANOVA on rapport in task 1 revealed no effects of the manipulated variables (as they had not yet occurred; Fs < 0.36, ps > .55), again confirming the success of random assignment. However, there was again a significant effect of culture ($F(1,190) = 12.26$, $p = .001$). Japanese also reported feeling less rapport than Americans (see Fig. 4).

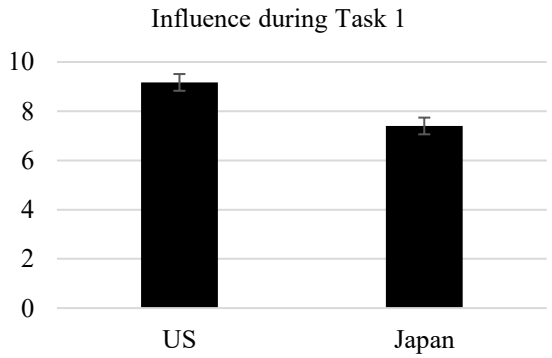

Figure 3: *Effect of culture on influence in task 1.*

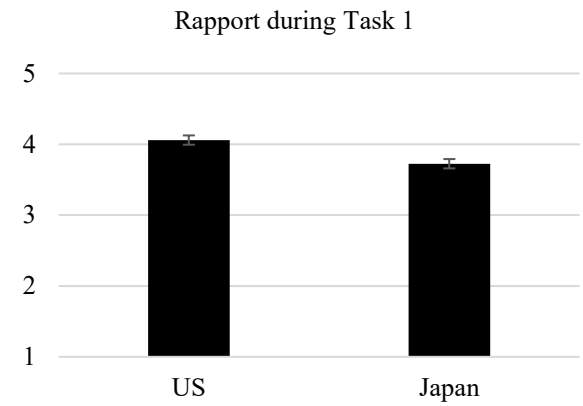

Figure 4: *Effect of culture on felt rapport in task 1.*

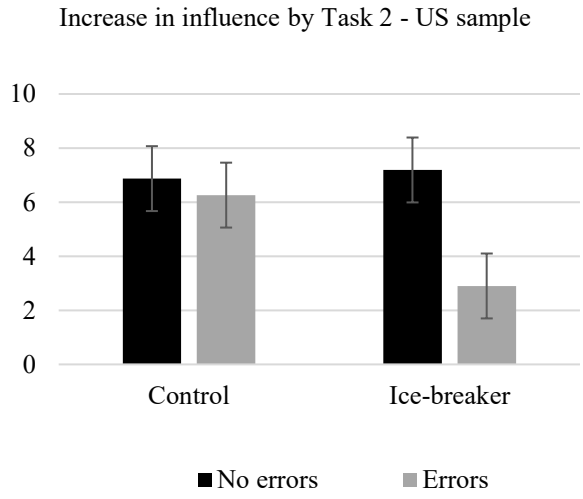

Figure 5: *Increase in influence from task 1 to 2 in U.S. sample.*

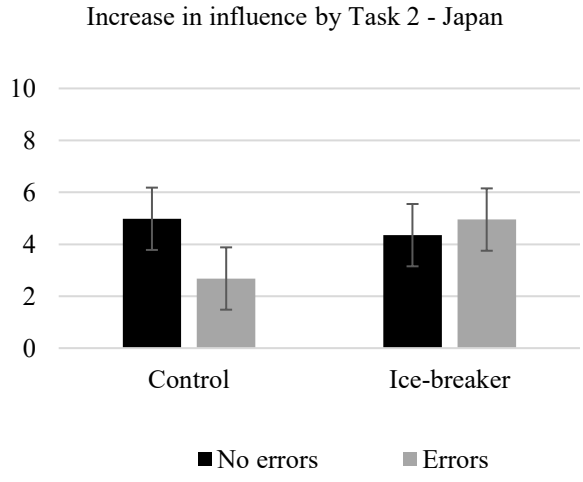

Figure 6: *Increase in influence from task 1 to 2 in Japan sample.*

4.2 Increase in influence by Task 2

Influence on the Save-the-Art Task was calculated, and then influence on the Lunar Survival task was subtracted from that score. This indexed increases in influence that occurred from the first to the second task. Increase in influence was tested using an ANCOVA, controlling for initial agreement on both tasks. There was a marginally significant effect of culture such that Japanese increased their influence over the course of the study less than Americans ($F(1,188) = 3.02$, $p = .08$). Accordingly, not only were the Japanese less influenced than Americans on the first task, their level of influence also increased (marginally) less from task 1 to task 2 compared to their U.S. counterparts.

There is also a marginally significant effect of errors, such that errors reduced the increase in influence compared to no errors ($F(1,188) = 3.64$, $p = .058$). Importantly, the effect of errors depended on both ice-breaker condition and culture, with a marginally significant three-way interaction ($F(1,188) = 3.62$, $p = .059$). As depicted in Fig. 5, the ice-breaker backfired among U.S. participants, reducing increases in influence after errors. That is, for the U.S. sample, errors were harmful to influence only when they occurred after the ice-breaker. However, as depicted in Fig. 6, errors reduced increases in influence in the absence of the ice-breaker among Japanese participants. Thus, the ice-breaker helped mitigate the influence of errors for the Japanese sample. All other main effects and interactions failed to approach significance (Fs < 1.82, ps > .18).

4.3 Increase in rapport by Task 2

Rapport felt during the Lunar Survival task was subtracted from rapport felt during the Save-the-Art Task. This indexed increases in rapport that occurred from the first to the second task, which was tested using an ANOVA. There was a significant effect of culture such that Japanese decreased their rapport over the course of the study more than Americans ($F(1,190) = 4.23$, $p = .04$; Fig. 7). There is also a significant effect of errors, such that errors led to reductions in rapport, whereas there were increase in rapport in the absence of errors ($F(1,190) = 50.11$, $p < .001$; Fig. 8). All other main effects and interactions failed to approach significance (Fs < 2.42, ps > .12).

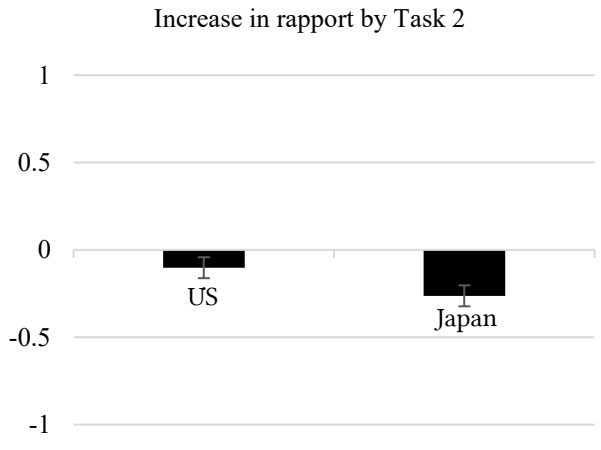

Figure 7: *Effect of culture on increase in rapport from task 1 to 2.*

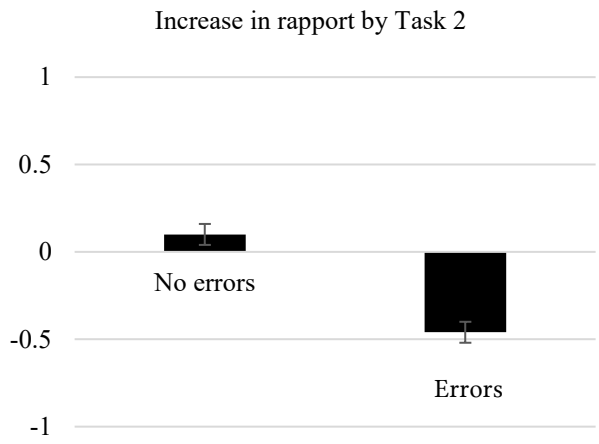

Figure 8: *Effect of errors on increase in rapport from task 1 to 2.*

5 DISCUSSION

5.1 Discussion of findings

In our study, Japanese participants were found to be less persuaded by the agent and reported feeling less rapport, as well as a greater reduction in rapport across the session. The latter could have occurred due to response bias. Culture-based response bias occurs when people of one culture distort responses to rating scales more than others. In Asian cultures, this often takes the form of selecting more moderate answers [41-44]. However, the Japanese participants also were less influenced by the agent, and showed reduced increase in influence over the course of the study, from the first influence task to the second. As this is a behavioral measure, it would not be due to response bias. Instead, perhaps Japanese are more frequently exposed to intelligent agents, especially those physically embodied as robots, and therefore any impact of novelty would be reduced for this sub-sample.

Besides considering culture, we also investigated the effect of conversational errors on influence and rapport. The presence of errors was marginally significantly detrimental to influence and statistically significantly detrimental to rapport. This replicates other work where agents that make conversational errors were found to be less capable of influencing people than agents that do not make errors. This occurs with both virtually represented agents [3,4] and physically embodied ones [33-34].

Perhaps more interesting than these main effects of culture and errors, we also found that culture interacted with the effects of errors and social dialogue. Recall that we sought to explore whether culture impacts the extent to which social dialogue can mitigate (or exacerbate) the loss of trust caused when agents make conversational errors. We observed a marginally significant three-way interaction between culture, errors and social dialogue.

First, as reported in [2], the pattern for U.S. participants showed that, among this cultural group, errors were only harmful to influence when the agent and user had previously engaged in an ice-breaker. That is, the effects of errors were driven by the ice-breaker condition. For Americans, this initial social dialogue seems to be required for subsequent errors to hinder influence.

This may be because the social dialogue effectively extends the agent's period of good performance. In the social dialogue condition, the agent performs well in the first task and then the ice-breaker. Without the social dialogue, the agent only performs well in the first task, so errors may seem less abrupt. It is possible, then, that, a contrast effect occurred for U.S. participants, whereby errors stood out more when the ice-breaker conversation occurred between the two tasks. The *contrast effect* is a common concept in social psychology, which notes that if someone has experienced a positive interaction, their response scale is anchored in the positive, and subsequent negative experiences are judged against that positive scale [45]. In the present study, our attempted rapport-building interaction backfired *if* it was followed by errors because that social dialogue extended the period of good performance before the errors, creating a contrast effect with the subsequent errors.

However, for the Japanese participants, this same social dialogue helps the agent to mitigate the detrimental impact of errors. Indeed, errors only hurt influence for Japanese participants in the absence of this social dialogue. Therefore, they did not exhibit a backfire effect of social dialogue like American participants did; instead, there was a benefit of engaging in the such a social interaction with our agent.

This marginally significant cross-cultural difference could have occurred for a number of reasons. First, it is possible that the contrast effect mechanism, or comparison to period of good performance, is present for Americans but not for Japanese. Second, perhaps a different kind of expectancy-violation occurred in the two different cultures. For Americans, the agent violated their expectations when it performed well, and then appeared to stop working during the second task. In contrast, for Japanese, the agent might have violated their expectations by not performing any kind of social dialogue (especially when reminded that this should occur by the control task, where a disembodied voice asked personal questions instead of the agent). In either case, this kind of expectancy-violation could account for the reduction in influence: in the American sample, it would have occurred in the social dialogue condition, but in the Japanese sample, it would have occurred in the control condition. This fits the observed pattern (depicted in Fig. 5 & Fig. 6).

There are other possible explanations for this apparent cross-cultural divergence, but some of them might not be as well-supported by the current data. For example, conversational agents are more common in Japan, making them more likely to have built some kind of connection to an agent before. However, prior cross-cultural work suggests that the stereotype that Japanese are thus more accepting of agents may not hold up. For example, Japanese were found to have less positive attitudes towards Aibo robot than U.S. users [46]. Yet, the Japanese's response may depend on the behavior of the agent. Because Japanese are more interdependent, being more liable to seek social connection [47], they may show more positive attitudes towards agents that behave more interdependently (like those that use social dialogue, as in our study). Indeed, research has found that members of more interdependent/collectivistic cultures rate more interdependent/collectivistic agents as more appropriate [48].

In our study, the Japanese participants might have viewed the exchange as more relational than the Americans (again because the former are more interdependent than the latter). Because of

this, the Japanese might have benefited more from the social dialogue. In contrast, without this orientation, the Americans just experienced the agent as a broken tool when errors occurred, and after a longer period of good performance (afforded by the ice-breaker), they were more frustrated and thus less influenced. Although this seems like a plausible account, the Japanese did not feel more connection with the agent (as they reported lower rapport). However, this interpretation may still be applicable. First, as mentioned above, the Japanese may only have reported lower rapport than the Americans because of a response bias. Additionally, the Japanese's higher expectations for connection to the agent may not only account for the observed pattern (as just described), but also account for the lower rapport ratings themselves (because lower expectations made the connection they had seem less impressive). Furthermore, the Japanese rapport ratings also decreased more over the course of the session than Americans'. This also could have been due to high expectations, which continued to go unmet by our agent.

Finally, while we attempted to keep the experimental procedure and conditions as identical across the American and Japanese samples, it is still possible that differences in interpretation between the two cultures could account for our findings. For example, particular information exchanges, such as attempts at persuasion, may carry different import in the two cultures. Japanese may have been less influenced (and more helped by social dialogue) because they cared less about success on a task that involved persuasion. Alternatively, the same type of error (e.g., saying that "I can't hear you") may be interpreted differently in the U.S. and Japan. These kinds of explanations may be less theoretically interesting than those posited above, however, they may still have contributed to the observed pattern of results.

5.2 Implications for design

Good performance in the form of social dialogue has been shown here to interact with culture and conversational errors. These results have implications for HCI and agent design. Designers should consider that, for users across cultures, conversational errors that occur in a social agent's dialogue hinder users from taking the agent's advice, undermining its persuasiveness.

Designers will reduce errors for a task to the best of their ability, given the available resources. Instead, our work takes on the question of deciding whether a social dialogue can help mitigate those errors. However, the benefit of including social dialogue such as an ice-breaker may depend on the users' culture. For Americans, errors seem to be particularly damaging when they suddenly appear after sustained good performance (here during a social dialogue). Designers should therefore consider alternative interventions other than including a social dialogue that extends the period of good performance. For American users, research could explore specific error-mitigation rapport-building, such as incorporating apologies, explanations, and negative self-disclosure. If the system is intended to influence American users, our results suggest it might be more effective to use another solution instead of trying to mitigate errors by preempting the impact of errors with a social dialogue. Research should further test recommendations.

These recommendations generalize, but only so far. For example, other work in independent cultures (e.g., western Europe) has found similar results. For example, it was found that an agent that changed from alignment to non-alignment during the course of a structured game was rated worse than an agent who improved to aligning after a first phase of non-alignment [49]. While this non-alignment was not an explicit conversational error, the results did, however, represent a similar interaction between conversational performance (good vs. bad) and the time dynamics of conversation. Hence, it seems that the results we found with the Americans may generalize to other independent cultures. However, from our results, they do not seem to generalize to interdependent cultures, at least not Japan.

Instead, in Japan, social dialogue may very well be a reasonable option for mitigating the impact of errors. In fact, our results suggest that, without this dialogue, the errors are detrimental to influence for the Japanese, but with the social dialogue, they are not. Designers working on agents for Japanese users (and perhaps members of other interdependent cultures) could consider using social dialogues to repair errors, even given periods of good performance. Further research should also consider cultures that lie on the spectrum between the most independent (e.g., U.S.) and most interdependent (e.g., Japan) cultures [47].

As a final note, given that one of our goals was to aid designers in deciding *when* a new (hence, error-prone) task should be placed in a set of established tasks, we fixed the ordering of task 1 and task 2 in this study. The Lunar Survival task always came before the Save-the-Art task. While there would have been some benefits to counterbalancing the tasks, we were theoretically interested in the question of *when* an error-prone task should be placed in an existing set of tasks. Additionally, statistical power would have been significantly diminished by having to add order as another between-subjects factor. Furthermore, as noted above, previous work suggests that that task order of these kinds of survival tasks would not affect results significantly [39]. Finally, this set task order cannot account for the observed differences between conditions, as participants of both cultures in all conditions completed the tasks in the same order. So the current study examines differences between cultures in how the various conditions impact the agent's ability to maintain influence across the tasks in this set order.

5.3 Conclusion

While errors reduce agents' influence marginally significantly across cultures, there were also marginal cultural differences in the likelihood that a rapport-building social dialogue can mitigate this detrimental impact of errors. For Japanese, the impact of errors only occurred in the absence of this social dialogue, and the errors did no harm if the agent engaged in the social dialogue with the user first. For Americans, instead of helping, the social dialogue appeared to exacerbate the problem by highlighting how well the agent functioned before the errors. While design could focus on other ways of mitigating errors for Americans, placing a social dialogue before the errors appears to be a poor option for persuasive agents designed for use in this culture. Future research should consider how other cultures respond to such design choices, including the role of social dialogue in overcoming errors.

ACKNOWLEDGMENTS

This work was supported in part by the U.S. Army. Any opinion, content or information presented does not necessarily reflect the position or the policy of the United States Government, and no official endorsement should be inferred.

REFERENCES

[1] Gratch, J., Wang, N., Gerten, J., Fast, E., & Duffy, R. Creating rapport with virtual agents. In International Workshop on Intelligent Virtual Agents, 125-138 (Springer, 2007).

[2] Lucas, G. M., Boberg, J., Artstein, R., Traum, D., Gratch, J., Gainer, A., Johnson, E., Leuski, A., & Nakano, M. Getting to know each other: The role of social dialogue in recovery from errors in social robots. Proceedings of the Conference on Human-Robot Interaction, 344-351 (2018).

[3] Wang, Y., Khooshabeh, P. & Gratch, J. Looking real and making mistakes. in *International Workshop on Intelligent Virtual Agents*, 339–348, (Springer, 2013).

[4] Blascovich, J. & McCall, C. Social Influence in Virtual Environments. in *The Oxford handbook of media psychology*, 305–315 (Oxford University Press, 2013).

[5] Wiegmann, D. A., Rich, A. & Zhang, H. Automated diagnostic aids: The effects of aid reliability on users' trust and reliance. *Theor. Issues Ergon. Sci., 2*, 352–367 (2001).

[6] Cassell, J. & Bickmore, T. Negotiated Collusion: Modeling Social Language and its Relationship Effects in Intelligent Agents. *User Model. User-adapt. Interact., 13*, 89–132 (2003).

[7] Traum, D., Swartout, W., Marsella, S. & Gratch, J. Fight, flight, or negotiate: Believable strategies for conversing under crisis. *Proc. Intell. Virtual Agents,* 52–64 (2005).

[8] Manuvinakurike, R., Velicer, W. F. & Bickmore, T. W. Automated indexing of Internet stories for health behavior change: weight loss attitude pilot study. *J. Med. Internet Res.,* 16, e285 (2014).

[9] Hiraoka, T., Neubig, G., Sakti, S., Toda, T. & Nakamura, S. Reinforcement Learning of Cooperative Persuasive Dialogue Policies using Framing. *In Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers* , 1706–1717 (2014).

[10] Hancock, P.A., Billings, D.R., Schaefer, K.E., Chen, J.Y., De Visser, E.J., & Parasuraman, R. A Meta-Analysis of Factors Affecting Trust in Human-Robot Interaction. *J. Hum. Factors Ergon. Soc., 53*, 517–527 (2011).

[11] Salem, M., Lakatos, G., Amirabdollahian, F. & Dautenhahn, K. Would You Trust a (Faulty) Robot? in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction - HRI '15*, 141–148 (ACM Press, 2015).

[12] Ham, J. & Midden, C. J. H. A persuasive robot to stimulate energy conservation: the influence of positive and negative social feedback and task similarity on energy-consumption behavior. *Int. J. Soc. Robot., 6,* 163–171 (2014).

[13] Khashe, S., Lucas, G. M., Becerik-Gerber, B., & Gratch, J. Buildings with persona: Towards effective building-occupant communication. Computers in Human Behavior, 75, 607-618 (2017).

[14] Albaina, I. M., Visser, T., van der Mast, C. A., & Vastenburg, M. H. Flowie: A persuasive virtual coach to motivate elderly individuals to walk. In G. Troester, K. Connelly, & B. Arnrich (Eds.), *Proceedings of the 3d International ICST Conference on Pervasive Computing Technologies for Healthcare.* ICST. (2009).

[15] Fox, J., & Bailenson, J. N. Virtual Self-Modeling: The Effects of Vicarious Reinforcement and Identification on Exercise Behaviors. *Media Psychology, 12(1),* 1–25. (2009).

[16] IJsselsteijn, W. A., Kort, Y. de, Bonants, R., Jager, M. de, & Westerink, J. Virtual Cycling: Effects of immersion and a virtual coach on motivation and presence in a home fitness application. *In Proceedings Virtual Reality Design and Evaluation Workshop*, 22–29 (2004).

[17] King, A. C., Bickmore, T. W., Campero, M. I., Pruitt, L. A., & Yin, J. L. Employing virtual advisors in preventive care for underserved communities: Results from the COMPASS study. *Journal of health communication, 18(12),* 1449–1464 (2013).

[18] Buttussi, F., Chittaro, L., & Nadalutti, D. Bringing mobile guides and fitness activities together: a solution based on an embodied virtual trainer. In *Proceedings of the 8th conference on Human-computer interaction with mobile devices and services,* 29-36 (ACM, 2006).

[19] Krämer, N., Karacora, B., Lucas, G. M., Dehghani, M., Rüther, G., & Gratch, J. Closing the gender gap in STEM with friendly male instructors? On the effects of rapport behavior and gender of a virtual agent in an instructional interaction. *Comput. Educ., 99,* 1–13 (2016).

[20] Tickle-Degnen, L. & Rosenthal, R. The nature of rapport and its nonverbal correlates. *Psychol. Inq., 1,* 285–293 (1990).

[21] Gremler, D. D. & Gwinner, K. P. Customer-Employee Rapport in Service Relationships. *J. Serv. Res., 3,* 82–104 (2000).

[22] Gfeller, J. D., Lynn, S. J. & Pribble, W. E. Enhancing hypnotic susceptibility: Interpersonal and rapport factors. *J. Pers. Soc. Psychol., 52,* 586 (1987).

[23] Carey, J. C., Stanley, D. A. & Biggers, J. Peak alert time and rapport between residence hall roommates. *J. Coll. Stud. Dev.* (1988).

[24] Fuchs, D. Examiner Familiarity Effects on Test Performance. *Topics Early Child. Spec. Educ., 7*, 90–104 (1987).

[25] Burns, M. Rapport and relationships: The basis of child care. *J. Child Care* (1984).

[26] Mutlu, B. & Forlizzi, J. Robots in organizations. in *Proceedings of the 3rd international conference on Human robot interaction - HRI '08*, 287 (ACM Press, 2008).

[27] Azhar, M. Q. Toward an argumentation-based dialogue framework for human-robot collaboration. in *Proceedings of the 14th ACM international conference on Multimodal interaction*, 305 (ACM Press, 2012).

[28] Clear, T., Daniels, M., Clear, T. & Daniels, M. A cyber-icebreaker for an effective virtual group? in *Proceedings of the 6th annual conference on Innovation and technology in computer science education, 33*, 121–124 (ACM Press, 2001).

[29] Mitchell, S. K. & Shepard M. Building Social Presence through Engaging Online Instructional Strategies. in *Student-Teacher Interaction in Online Learning Environments*, 133–156 (2015).

[30] Risvik, L. L. Framework for the Design and Evaluation of Icebreakers in Group Work. Master's Thesis. (2014).

[31] Sidner, C. L., Kidd, C. D., Lee, C. & Lesh, N. Where to look: a study of human-robot engagement. in *Proceedings of the 9th international conference on Intelligent user interfaces*, 78–84 (ACM, 2004).

[32] Gockley, R., Bruce, A., Forlizzi, J., Michalowski, M., Mundell, A., Rosenthal, S., Sellner, B., Simmons, R., Snipes, K., Schultz, A.C., & Wang, J. Designing robots for long-term social interaction. in IEEE/RSJ *International Conference on Intelligent Robots and Systems*, 1338–1343 (IEEE, 2005).

[33] Desai, M., Medvedev, M., Vázquez, M., McSheehy, S., Gadea-Omelchenko, S., Bruggeman, C., Steinfeld, A., & Yanco, H. Effects of changing reliability on trust of robot systems. in *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction - HRI '12 73* (ACM Press, 2012).

[34] Desai, M., Kaniarasu, P., Medvedev, M., Steinfeld, A. & Yanco, H. Impact of robot failures and feedback on real-time trust. *Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction,* 251–258 (2013).

[35] Moon, Y. & Youngme. The effects of distance in local versus remote human-computer interaction. in *Proceedings of the SIGCHI conference on Human factors in computing systems*, 103–108 (ACM Press, 1998).

[36] Khooshabeh, P., McCall, C., Gandhe, S., Gratch, J. & Blascovich, J. Does it matter if a computer jokes. in *CHI'11 Extended Abstracts on Human Factors in Computing Systems,* 77–86 (ACM, 2011).

[37] Ramachandran, D. & Canny, J. The persuasive power of human-machine dialogue. in *International Conference on Persuasive Technology*, 189–200 (Springer, 2008).

[38] Kelley, J. F. An iterative design methodology for user-friendly natural language office information applications. *ACM Trans. Inf. Syst.*, 2, 26–41 (1984).

[39] Artstein, R., Traum, D., Boberg, J., Gainer, A., Gratch, J., Johnson, E., Leuski, A., & Nakano, M. Listen to My Body: Does Making Friends Help Influence People? *Proceedings of the Thirtieth International Florida Artificial Intelligence Research Society Conference,* 430-435. (2017).

[40] Von der Pütten, A. M., Krämer, N. C., Gratch, J. & Kang, S.-H. 'It doesn't matter what you are!' Explaining social effects of agents and avatars. *Comput. Human Behav., 26*, 1641–1650 (2010).

[41] Byrne, B. M., & Campbell, T. L. Cross-cultural comparisons and the presumptions of equivalent measurement and theoretical structure: A look beneath the surface. *Journal of Cross-Cultural Psychology, 30,* 555-574. (1999).

[42] Hofstede, G. *Culture's consequences: International differences in work-related values.* Beverly Hills, CA: Sage. (1980).

[43] Fischer, R. Standardization to account for cross-cultural response bias: A classification of score adjustment procedures and review of research in JCCP. *Journal of Cross-Cultural Psychology, 35(3),* 263-282. (2004).

[44] Steele, C. M., & Aronson, J. Stereotype threat and the intellectual test performance of African Americans. *Journal of personality and social psychology, 69(5),* 797. (1995).

[45] Sherman, S. J., Ahlm, K., Berman, L. & Lynn, S. Contrast effects and their relationship to subsequent behavior. *J. Exp. Soc. Psychol., 14,* 340–350 (1978).

[46] Bartneck, C., Suzuki, T., Kanda, T., & Nomura, T. The influence of people's culture and prior experiences with Aibo on their attitude towards robots. *Ai & Society, 21(1-2),* 217-230 (2007).

[47] Markus, H. R., & Kitayama, S. Culture and the self: Implications for cognition, emotion, and motivation. *Psychological review, 98(2),* 224. (1991).

[48] Mascarenhas, S., Degens, N., Paiva, A., Prada, R., Hofstede, G. J., Beulens, A., & Aylett, R. Modeling culture in intelligent virtual agents. *Proceedings of Autonomous Agents and Multi-Agent Systems, 30(5),* 931-962 (2016).

[49] Nuñez, T. R., Prynda, K., Bergmann, & K. Von der Pütten, A. M. Computers Aligning to Their Users: Lexical Alignment in Human-Agent-Interaction and Its Psychological Effects. *Virtual Social Interaction Workshop.* (2017).