"We Need to Start Thinking Ahead": The Impact of Social Context on Linguistic Norm Adherence

Jane Lockshin (jlockshin@mines.edu)
MIRRORLab, 1600 Illinois Street
Golden, CO 80401 USA

Tom Williams (twilliams@mines.edu)
MIRRORLab, 1600 Illinois Street
Golden, CO 80401 USA

Abstract

Human dialogue is governed by communicative norms that speakers are expected to follow in order to be viewed as cooperative dialogue partners. Accordingly, for language-capable autonomous agents to be effective human teammates they must be able to understand and generate language that complies with those norms. Moreover, these linguistic norms are highly context sensitive, requiring autonomous agents to be able to model the contextual factors that dictate when and how those norms are applied. In this work, we consider three key linguistic norms (directness, brevity, and politeness), and examine the extent to which adherence to these norms varies under changes to three key contextual factors (potential for harm, interlocutor authority, and time pressure). Our results, based on a human-subject study involving 5,642 human utterances, provide strong evidence that speakers do indeed vary their adherence to these norms under changes to these contextual factors.

Keywords: Learning Human Values and Preferences; Linguistic Norms; Human-Robot Interaction

Introduction

Human dialogue is governed by communicative norms that speakers are expected to follow in order to be viewed as cooperative dialogue partners. Grice, for example, delineates four conversational maxims (Quantity, Quality, Relation, and Manner), which he stipulates that humans automatically assume will be followed by all speakers, according to a general *Cooperativity Principle:* "Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged."(Grice, 1975). Accordingly, for language-capable autonomous agents to be effective human teammates they must be able to understand and generate language that complies with those norms. Indeed, research has demonstrated that robots, chatbots, and other autonomous agents that uniformly fail to comply with Gricean maxims are viewed as less humanlike (Baratgin, Jacquet, & Cergy, 2019; Saygin & Cicekli, 2002; Jacquet, Baratgin, & Jamet, 2018).

Humans do not, however, always comply with these norms themselves – in fact, it is well understood that humans regularly flout these norms in order to satisfy other communicative goals or to adhere to other sociocultural norms. For example, humans regularly violate these norms for the purpose of *implicature*, i.e., to use the very violation of these norms to convey additional information, especially information that cannot otherwise be securely or tactfully communicated. Researchers in Human-Robot Interaction (HRI) have demonstrated the importance of conversational implicature for artificial agents designed to interact with robots, showing that artificial agents can attain higher task success when they strategically violate norms for the sake of implicature (Liang, Proft, Andersen, & Knepper, 2019) – and fail to understand most of what is said to them when they are unable to process common forms of conversational implicature such as Indirect Speech Acts (Williams, Thames, Novakoff, & Scheutz, 2018), which are typically employed to comply with *other* sociocultural norms, related to politeness.

Whether or not humans choose to comply with linguistic and sociocultural norms is highly context-sensitive. By speaking indirectly, speakers choose to violate Grice's Maxim of Manner ("Be perspicuous") in order to avoid violating sociocultural politeness norms. Accordingly, ones' willingness to violate Grice's Maxim of Manner should be sensitive to the extent to which their context demands politeness. Individuals in a workplace may speak to their colleagues differently than their clients, for example, and may speak to their clients differently when they are involved in solving time-sensitive issues than when they are engaging in routine tasks. Indeed, previous work has demonstrated exactly this sort of context sensitivity, both in human-human dialogue (Agha, 2006) and human-robot dialogue (Williams et al., 2018) and has demonstrated the benefits of robots appropriately appropriately adapting their utterances to changing social and conversational contexts (Ritschel, Baur, & André, 2017; Jackson, Wen, & Williams, 2019).

The previously discussed bodies of work have clearly demonstrated that robots must be able to understand and generate language that complies with linguistic norms, including broadly applicable Gricean Maxims and context-sensitive sociocultural norms, and as such, must be able to understand the relationship between specific contexts and the context-sensitive norms that should be adhered to within that context. To address this challenge, Gervits, Briggs, and Scheutz (2017) presented a ranking algorithm that selects between different utterance phrasings based on context-sensitive priority rankings over linguistic norms. A significant limitation of this approach, however, is that Gervits et al.'s rankings are directly associated with high-level contexts, such as service operations, military missions, and home emergencies. We believe that there are three primary shortcomings to this approach: (1) it does not appropriately capture *what it*

is about that context that produces the associated norm prioritization ranking; and because of this, (2) their approach is unable to vary how norms may need to be selectively employed *within* high level contexts, and (3) it may be difficult to extend their approach to new contexts.

We argue that the ideal approach to context-sensitive robot language understanding and generation instead requires understanding not of relationships between norms and high-level contexts, but rather of relationships between norms and *contextual factors*; i.e., the features of those contexts that are actually responsible for the need for those norms to be followed (or not). Specifically, we argue for a model in which norm adherence can be expressed as a network of the form:

$$P(N,I,F,C) = \prod_{N_i \in N} P(N_i|I)P(I) \prod_{F_j \in F} P(N_i|F_j)P(F_j|C)P(C)$$

That is, adherence to each norm N_i in norm set N depends on (1) certain properties of the intention itself I, and (2) independent from the intention to be communicated, a set of contextual factors F, which are themselves informed by the current context C.

This model thus requires three components to be learned: (1) the relationship between individual norms (e.g., directness, politeness, brevity) and properties of the intention to be communicated (e.g., a simple acknowledgement is perhaps a priori more likely to be conveyed briefly than is a request for information); (2) the relationship between individual norms and individual contextual factors (e.g., utterances issued under time pressure are perhaps more likely to be conveyed briefly than utterances not issued under time pressure); and (3) the relationship between contextual factors and high-level contexts (e.g., time pressure is more likely during a task execution than during post-task debriefing).

In this paper, we explore the second of these relationships (i.e., $P(N_i|F_j)$); specifically, (1) what linguistic norms are sensitive to context? (2) what specific contextual factors do we expect to impact those linguistic norms? And (3) how can the relationship between norm adherence and contextual factors be learned from human-human interactions?

The rest of the paper proceeds as follows: in the next section, we formally define the linguistic norms and contextual factors we are interested in investigating in this work. Then, we present the results of a human-subject experiment that allows us to analyze the relationship between contextual factors and linguistic norm adherence. Finally, we discuss how our findings may be applied in the future to context-sensitive language generation for interactive robots.

Key Features

To introduce our approach, we first formally define our linguistic norms and contextual factors of interest.

Linguistic Norms

Linguistic norms are the loosely-defined "rules" that are adhered to by conversational partners in order to effectively communicate their intentions while maintaining sensitivity to their social context (Roughley & Bayertz, 2019), and include the use of idiomatic language and cultural conventions that govern language. In this work, we are interested in examining adherence to the linguistic norms of directness, brevity, and politeness (i.e., the same linguistic norms examined by Gervits et al. (2017)), given the presence of one or more contextual factors.

Directness – Human interlocutors use conventionally indirect forms such as "Could you X" or "Would you mind X" to avoid being perceived as impolite. This is especially true in the case of indirect requests, where a direct command would be viewed as requiring the listener to perform the speaker's desired action at the expense of achieving their own implicitly desired actions. Speaking in this way, however, violates Grice's (a portion of) Grice's Maxim of Manner (i.e., to avoid speaking ambiguously or unclearly). Yoon, Tessler, Goodman, and Frank (2016) refer to this as a tradeoff between epistemic and social utility. We classify an utterance as *direct* if it does not take the form of a conventionalized *indirect speech act*, in which the literal and intended meanings noticeably differ in a way that is standardized according to cultural convention (Searle, 1975).

Brevity – Human interlocutors are brief in order to adhere to (a portion of) Grice's Maxim of Manner (namely, "Be brief") (although one could argue that brevity in some circumstances could avoid negative face threat (Brown, Levinson, & Levinson, 1987) by avoiding imposing on the listener's time). We classify an utterance as *brief* when the number of words in that utterance falls below some threshold τ.

Politeness – Strategies for mitigating utterance face-threat vary across social relationships and sociocultural contexts (Haugh & Chang, 2015), but often include gratitude, deference, apologizing, and saying "please" (Danescu-Niculescu-Mizil, Sudhof, Jurafsky, Leskovec, & Potts, 2013) (as well as the use of indirect speech acts (Clark & Schunk, 1980)). These strategies typically avoid violation of sociocultural politeness norms while violating Grice's Maxim of Manner (by, again, failing to be brief). We classify an utterance as *polite* if it exhibits any of these politeness strategies (other than indirectness, due to our interest in capturing indirectness in and of itself) (Brown et al., 1987).

Contextual Factors

While researchers such as Yoon, Tessler, Goodman, and Frank (2017) emphasize the role of individual speakers' personal trait weightings on epistemic vs. social utilities, we instead focus on the role played by interactants' shared state context. The *contextual factors* that define this state context alter the social climate, and accordingly, the norms that agents are expected to adhere to in that social context. The three contextual factors that we are interested in examining in this work are: potential for harm, interlocutor authority, and time pressure. These contextual factors were chosen based on their prevalence and variability across a variety of social domains.

Potential for harm occurs when a critical situation presents a serious possibility for negative outcomes for agents within a context if the correct actions are not carefully executed. For example, an alarm that signals an emergency inside of a building signifies potential for harm for the individuals located in that building. When individuals take necessary actions and precautions to alleviate the emergency that triggered the alarm, potential for harm is reduced.

Interlocutor authority occurs when the agent who is being spoken to possesses authority over the agent who is speaking, or vice versa. For example, restaurant managers possess organizational authority over their employees, and paying customers of the restaurant may also have some degree of perceived social authority over the restaurant workers.

Time pressure occurs when there is a limited amount of time to complete a task. For example, if an individual is given five minutes to solve a puzzle, time pressure exists during these five minutes (and disappears after the allocated time concludes).

Method

We conducted a human-subjects experiment to examine how adherence to our norms of interest varied according to the presence/absence of our contextual factors of interest. This experiment received ethics approval from our Human-Subjects Research office.

Experimental Design

We identified the board game *Pandemic*[1] as an ideal example of a context in which our contextual factors of interest could be systematically varied (Leacock, 2018). Pandemic is a cooperative board game in which players work with one another as staff members of the Center for Disease Control in order to cure four diseases that are spreading around the globe. Pandemic was chosen due to the game's collaborative nature (all players win or lose together, so communication between players is required), and because it allows for systematic variation of the contextual factors of interest (potential for harm, interlocutor authority, and time pressure).

Throughout the game, Pandemic allows for systematic variation of potential for harm by simulating the infection and spreading of diseases. Potential for harm is considered to be active when a significant number of areas around the world are infected with diseases (indicating that the players are close to losing the game); a condition that is alleviated when those diseases are marked as cured.

Interlocutor authority is systematically varied as players take their turns throughout the game. On each player's turn, that player decides what action to take, but can solicit suggestions and engage in strategic discussion with their fellow teammates. Thus, the player whose turn it is holds authority over the other players.

Figure 1: Participants playing the board game *Pandemic* during a lab experiment.

Finally, to systematically vary time pressure in the game, we introduced a between-subjects experimental manipulation: while the game by default does not include any explicit time pressure, we introduced a *timed* version of the game in which participants were given 90 seconds to take their actions (including all collaborative discussion time needed to decide what those actions should be).

In Pandemic, each turn consists of taking 1-4 'actions' (moves performed each turn to cure or prevent the spread of diseases), drawing cards that are needed to cure diseases, and drawing cards that spread diseases. If a player draws an epidemic card, a new city becomes infected with a disease, and cities that are already infected increase their likelihood of spreading disease. To keep game difficulty and length consistent across games, each game session was played with four epidemic cards evenly spaced throughout the game's epidemic deck.

Procedure

Participants were recruited in groups to play Pandemic through web postings and flyers at the Colorado School of Mines. During time-slot session assignment, we ensured that members of each group had no prior social connection to one another.

Upon arrival, participants were given as much time as needed to read and sign consent forms, after which they received verbal and printed game instructions. Participants were told that they were allowed to refer to the printed instructions throughout the duration of the game, if needed.

After addressing any questions or concerns regarding the game rules, the experimenter left the experiment area, and video and audio recording of the experiment began. Participants were then left to play Pandemic until they either won or lost the game. Half of the experiment groups played the traditional untimed version of Pandemic, and the other half played our timed version.

Upon completion of the game, the experimenter debriefed the participants as to the true purpose of the study. Before

[1]This experiment was conducted in 2019, before the global outbreak of COVID-19. Researchers seeking to employ similar methodology in the future without the use of a Pandemic-themed setting may wish to instead use the board game "Forbidden Island", from the same game creator (Leacock, 2011).

Directness
Direct Utterances
"Build a research station in that city."
"Move to the nearest city containing the yellow disease."
"I put that card in the discard pile."
Indirect Utterances
"Can you build one there?"
"I think you should move here."
"You could have discarded that one, I believe."

Table 1: Examples of direct versus indirect speech acts observed in our experiment.

Brevity
Brief Utterances
"Move there."
"Hand me the blue."
"Pick up two."
Non-Brief Utterances
"Now move there and cure that disease."
"Are you able to hand me that card?"
"You must pick up two cards at the end of your turn."

Table 2: Examples of brief versus non-brief speech acts observed in our experiment.

departing the experiment area and receiving payment, participants were asked to fill out a demographic form to identify age, gender, and education level.

Participants

Data was collected from 18 participants (5 male, 13 female) across six three-participant games. Participant ages ranged from 20 to 33 years of age (M=27.11,SD=3.03). All participants were paid $5 for their participation. All participants were students from Colorado School of Mines.

Data Annotation

Videos of each experimental session were transcribed and annotated using the ELAN video annotation software (Borovanskỳ, Kirchner, Kirchner, Moreau, & Vittek, 1996), with annotations for the following events: beginning and end of players' turns; important game events (e.g., drawing of an epidemic card or curing a disease); and player utterances. Each player utterance was annotated with the following information: which player was speaking; the content of their utterance; and the linguistic norms followed by their utterance (directness, brevity, and/or politeness). We will now describe the criteria used to evaluate whether each linguistic norm was followed in a given utterance:

We categorized utterances as *direct* if and only if they did not match the form of common conventionalized indirect speech acts (Searle, 1975) found in Briggs, Williams, and Scheutz (2017)'s indirect speech act taxonomy. Examples of direct versus indirect speech acts observed in our experiment can be found in Table 1.

We categorized an utterance as *brief* if its length in words was within the bottom third of utterance lengths in the game in which it was verbalised. Examples of brief utterances can be found in Table 2

We categorized an utterance as *polite* if it reflected one or more of the following linguistic strategies: gratitude, deference, apologizing, compliments, and "please" (Danescu-Niculescu-Mizil et al., 2013). Examples of polite versus non-polite utterances can be found in Table 3.

After each utterance was identified as direct, brief, and/or polite, the annotated transcript was used to identify which

Politeness
Polite Utterances
"Yeah, smart of you for thinking a few moves ahead."
"You treated China, nice."
"Pass me the red card, please."
Non-Polite Utterances
"We need to start thinking ahead."
"We already cured that."
"Hand me that one."

Table 3: Examples of polite versus non-polite speech acts from our data.

contextual factors held when the speech act was uttered. This produced a dataset of utterances encoded with both the norms adhered to in their production and the contextual factors under which they were produced.

In order to identify which contextual factors were adhered to during the production of each utterance, the following elements were tracked in our video transcripts: game progress, participant turns, and whether or not participants were timed.

The presence of *potential for harm* is determined by certain events that alter the current progress of the game (when diseases spread or become cured), and the state of the board itself (markers on the board indicate whether or not the game is close to completion). Specifically, potential for harm exists if the following logical expression for harm (H) holds true:

$$H = M \vee P \vee R \vee ((E \vee O) \wedge \neg(C \vee D))$$

Where:

- M = Markers that track infections and outbreaks are at least 75% of the way to the end (the last placement for the markers signifies the end of the game).
- P = There are less than five cards in the player card deck (running out of cards in the player card deck signifies the end of the game).
- R = Less than five cubes (the "infections") of a disease remain off the board (if all cubes of one disease are infected on the board, the game ends).

- E = An epidemic card is drawn.
- O = An outbreak occurs (when a city has too much of a disease and infects all of its adjacent cities).
- C = No city has more than two cubes (diseases) of the same color.
- D = A disease is cured on the current turn.

Interlocutor authority is determined by whose turn it is in the game: the speaker has authority on their turn.

Finally, the presence of *time pressure* is determined by the running of the timer in the game. When the timer starts, all utterances spoken during the 90 seconds of time allocated to perform actions were considered to have time pressure, and the utterances spoken after the completion of the timer (during the drawing of cards) are not considered to have time pressure.

Analysis

The six experiments we conducted provided us with a total of 5,642 utterances. The annotated utterances are available at `https://osf.io/m92as/`. This dataset was analyzed using a logistic regression to ascertain the effects of potential for harm, interlocutor authority, and time pressure, on likelihood of participants' utterances exhibiting directness, brevity, and politeness-beyond-indirectness. This analysis was performed using the JASP statistical software package (JASP Team, 2019), and an α level of 0.05 was used to establish statistical significance.

Results

In this section we report each of our three analyses (i.e., the analyses for each of these three linguistic norms).

Directness: The logistic regression analysis with directness as the dependent variable was statistically significant ($\chi^2(5634) = 50.942$, $p < .001$). Specifically, as shown in the top two plots of Fig. 2, significant effects were found for potential for harm ($p = .001$) and of time pressure ($p < .001$): utterances delivered under potential for harm were 1.378 times more likely to be direct, and utterances delivered under time pressure were 1.486 times more likely to be direct.

Brevity: The logistic regression analysis with brevity as the dependent variable was not statistically significant ($\chi^2(5634) = 10.140$, $p = .181$).

Politeness: The logistic regression analysis with politeness as the dependent variable was statistically significant ($\chi^2(5634) = 68.861$, $p < .001$). Specifically, as shown in the bottom two plots of Fig. 2, significant effects were found for potential for harm ($p < .001$), time pressure ($p = .011$), and the interaction of potential for harm, time pressure, and interlocutor authority ($p = .004$): utterances delivered under potential for harm were 1.59 times less likely to be polite, utterances delivered under time pressure were 1.348 times less likely to be direct, but utterances delivered under potential for harm and time pressure while speaking to an authority were 2.721 times *more* likely to be polite.

Contextual Factor(s)	*Directness*	*Brevity*	*Politeness*
PFH, IA, TP	0.42	0.29	0.29
PFH, IA	0.43	0.30	0.26
PFH, TP	0.49	0.39	0.12
IA, TP	0.42	0.33	0.26
PFH	0.43	0.34	0.23
IA	0.34	0.33	0.33
TP	0.45	0.29	0.26
None	0.36	0.32	0.32

Table 4: Probabilities of applying the linguistic norms of directness, brevity, and politeness, given the presence of potential for harm (PFH), interlocutor authority (IA), and/or time pressure (TP).

Discussion

Our results suggest that contextual factors may be used to help predict adherence to sociocultural linguistic norms of directness and politeness. Specifically, our results illustrate that humans speak more directly when potential for harm and/or time pressure is present, and speak less politely when these factors are present *except* when speaking to someone in a position of authority.

We did not, however, find any ability for these contextual factors to predict adherence to brevity norms. This is most likely due to the disproportionate influence of utterance type on brevity; some utterance types (e.g., acknowledgements) are universally brief, while others may tend to be longer. In future work we plan to conduct additional analysis of our collected dataset to learn not only the $P(N_i|F_j)$ *model term learned in this work, but also the* $\mathrm{P}(\mathrm{N}_i|I)$ model term of the model defined in our introduction.

This work also provides progress towards our ultimate goal: building a predictive model that will allow a robotic agent to autonomously decide which norms to adhere to in different contextual conditions. As shown in Table 4, while our current analysis allows us to observe significant differences in norm adherence between different conditions, if used as a predictive model this would necessarily produce the same decision in all conditions (e.g., in all conditions the model would argue *against* adherence to politeness norms, albeit with different levels of confidence). Again, it is our hope that this shortcoming will be addressed when additional *non-context-dependent* factors inherent to the to-be-communicated intention are accounted for.

Another potential modification to our proposed predictive model would be to represent all model variables as continuous rather than discrete (Cobb, Rumí, & Salmerón, 2007): potential for harm can be assessed as a continuous score based on different harmful factors; time pressure can be assessed as the amount of time remaining; interlocutor authority can be assessed based on degrees of social distance; directness can be assessed based on level of conventional indirectness; politeness can be assessed based on number and type of

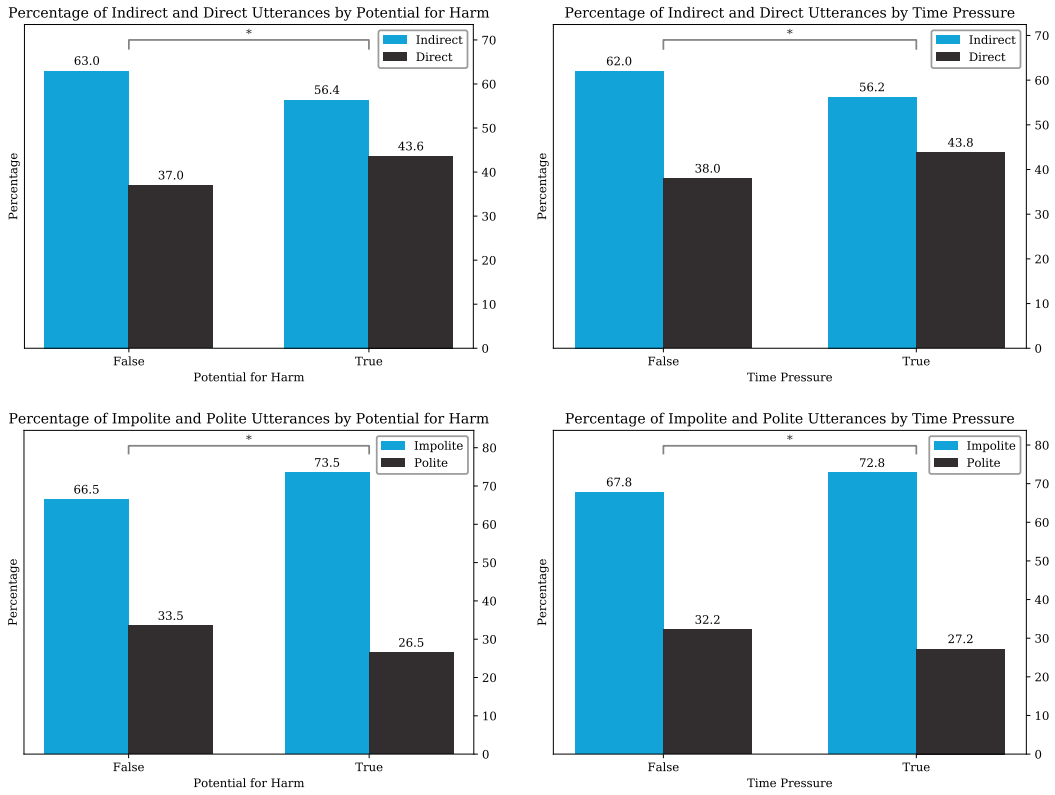

Figure 2: Effects of potential for harm (left) and time pressure (right) on directness (top) and politeness (bottom)

politeness markers employed; and brevity, if retained, can be assessed based on number of words used. Using continuous rather than discrete model variables would allow us to both perform more nuanced statistical testing and allow a predictive model to produce more nuanced results.

In future work, we also aim to integrate this proposed model into a cognitive robotic architecture, so that it can be leveraged to effectively generate contextually-appropriate robot language. Specifically, we plan to integrate our model into the ADE implementation (Kramer & Scheutz, 2006) of the Distributed, Integrated, Affect, Reflection, Cognition (DIARC) architecture (Scheutz et al., 2013, 2019). After completing this integration, we plan to empirically examine the effectiveness of our model in enabling more natural and contextually appropriate human-robot interactions and the learning of relationships between different contexts and the contextual factors explored in this paper, and to assess how this could enable robots to seamlessly adapt their language as they change between different real-world contexts.

Conclusion

Designing natural language-capable artificial agents that can communicate fluidly and appropriately across contexts requires defining features that allow these agents to abide by the same context-sensitive linguistic norms as their human teammates. Regardless of the overarching task context that these agents are designed for, the linguistic norms that should be adhered to will change along with changes in robots' environmental and social context (Malle, Scheutz, & Austerweil, 2017). To better understand the impact on these contextual factors on linguistic norm adherence, we thus examined the relationship between key linguistic norms (directness, brevity, and politeness) and the presence of potential for harm, interlocutor authority, and time pressure.

Evaluation of our collected data indicates that contextual factors play a significant role in determining politeness and directness, but further analysis needs to be conducted to examine why no such effect was observed on speaker brevity. In the future, we plan to leverage the results of this experiment to develop a Bayesian Network model of linguistic norm adherence integrated into a cognitive robot architecture, and conduct additional human-subjects experiments in our lab to assess this model. These experiments will allow us to further evaluate our approach's performance. Ultimately, we hope that our proposed model may enable intelligent agents to better engage in natural dialogue with their human teammates.

Acknowledgments

This work was supported in part by NSF Grant IIS-1849348.

References

Agha, A. (2006). *Language and social relations* (Vol. 24). Cambridge University Press.

Baratgin, J., Jacquet, B., & Cergy, F. (2019). Cooperation in online conversations: the response times as a window into the cognition of language processing. *Frontiers in psychology*, *10*, 727.

Borovanskỳ, P., Kirchner, C., Kirchner, H., Moreau, P.-E., & Vittek, M. (1996). Elan: A logical framework based on computational systems. *Electronic Notes in Theoretical Computer Science*, *4*, 35–50.

Briggs, G., Williams, T., & Scheutz, M. (2017). Enabling robots to understand indirect speech acts in task-based interactions. *Journal of Human-Robot Interaction*, *6*(1), 64–94.

Brown, P., Levinson, S. C., & Levinson, S. C. (1987). *Politeness: Some universals in language usage* (Vol. 4). Cambridge university press.

Clark, H. H., & Schunk, D. H. (1980). Polite responses to polite requests. *Cognition*, *8*(2), 111–143.

Cobb, B. R., Rumí, R., & Salmerón, A. (2007). Bayesian network models with discrete and continuous variables..

Danescu-Niculescu-Mizil, C., Sudhof, M., Jurafsky, D., Leskovec, J., & Potts, C. (2013). A computational approach to politeness with application to social factors. *arXiv preprint arXiv:1306.6078*.

Gervits, F., Briggs, G., & Scheutz, M. (2017). The pragmatic parliament: A framework for socially-appropriate utterance selection in artificial agents..

Grice, H. P. (1975). Logic and conversation. In *Speech acts* (pp. 41–58). Brill.

Haugh, M., & Chang, W.-L. M. (2015). Understanding im/politeness across cultures: an interactional approach to raising sociopragmatic awareness. *International Review of Applied Linguistics in Language Teaching*, *53*(4).

Jackson, R. B., Wen, R., & Williams, T. (2019). Tact in noncompliance: The need for pragmatically apt responses to unethical commands. In *Proceedings of the aaai/acm conference on artificial intelligence, ethics, and society.*

Jacquet, B., Baratgin, J., & Jamet, F. (2018). The gricean maxims of quantity and of relation in the turing test. In *2018 11th international conference on human system interaction (hsi)* (pp. 332–338).

JASP Team. (2019). *JASP (Version 0.10.2)[Computer software].* Retrieved from `https://jasp-stats.org/`

Kramer, J., & Scheutz, M. (2006). Ade: A framework for robust complex robotic architectures. In *2006 ieee/rsj international conference on intelligent robots and systems* (pp. 4576–4581).

Leacock, M. (2011). *Forbidden island.* `https://gamewright.com/product/Forbidden-Island`.

Leacock, M. (2018). *Pandemic.* `https://www.zmangames.com/en/games/pandemic/`.

Liang, C., Proft, J., Andersen, E., & Knepper, R. A. (2019). Implicit communication of actionable information in human-ai teams. In *Proceedings of the 2019 chi conference on human factors in computing systems* (pp. 1–13).

Malle, B. F., Scheutz, M., & Austerweil, J. L. (2017). Networks of social and moral norms in human and robot agents. In *A world with robots* (pp. 3–17). Springer.

Ritschel, H., Baur, T., & André, E. (2017). Adapting a robot's linguistic style based on socially-aware reinforcement learning. In *2017 26th ieee international symposium on robot and human interactive communication (ro-man)* (pp. 378–384).

Roughley, N., & Bayertz, K. (2019). Linguistic norms. *The Normative Animal?: On the Anthropological Significance of Social, Moral, and Linguistic Norms*, 38.

Saygin, A. P., & Cicekli, I. (2002). Pragmatics in human-computer conversations. *Journal of Pragmatics*, *34*(3), 227–258.

Scheutz, M., Briggs, G., Cantrell, R., Krause, E., Williams, T., & Veale, R. (2013). Novel mechanisms for natural human-robot interactions in the diarc architecture. In *Workshops at the twenty-seventh aaai conference on artificial intelligence.*

Scheutz, M., Williams, T., Krause, E., Oosterveld, B., Sarathy, V., & Frasca, T. (2019). An overview of the distributed integrated cognition affect and reflection diarc architecture. In *Cognitive architectures* (pp. 165–193). Springer.

Searle, J. R. (1975). Indirect speech acts. *Syntax & Semantics, 3: Speech Act*, 59–82.

Williams, T., Thames, D., Novakoff, J., & Scheutz, M. (2018). "thank you for sharing that interesting fact!": Effects of capability and context on indirect speech act use in task-based human-robot dialogue. In *Proceedings of the 13th acm/ieee international conference on human-robot interaction.*

Yoon, E. J., Tessler, M. H., Goodman, N. D., & Frank, M. C. (2016). Talking with tact: Polite language as a balance between kindness and informativity. In *Proceedings of the 38th annual conference of the cognitive science society* (pp. 2771–2776).

Yoon, E. J., Tessler, M. H., Goodman, N. D., & Frank, M. C. (2017). " i won't lie, it wasn't amazing": Modeling polite indirect speech. In *Cogsci.*