How to Tune your Draggin': Can Body Language Mitigate Face Threat in Robotic Noncompliance?*

Aidan Naughton and Tom Williams

Colorado School of Mines Department of Computer Science
Golden, CO 80401 {aidannaughton, twilliams}@mines.edu

Abstract. When social robots communicate moral norms, such as when rejecting inappropriate commands, humans expect them to do so with appropriate tact. Humans use a variety of strategies to carefully tune their harshness, including variations in phrasing and body language. In this work, we experimentally investigate how robots may similarly use variations in body language to complement changes in the phrasing of moral language.

Keywords: Human-robot communication. Social and moral norms.

1 INTRODUCTION

Robots are being increasingly used in more morally sensitive contexts such as healthcare, elder care, and military domains [35, 37, 2]. Because robots are perceived as moral and social agents, they are expected to adhere to the same moral norms that humans do. When robots fail to do so, negative attributions such as blame are often attached to the interaction [23]. Accordingly, researchers have argued that robots designed for morally sensitive contexts must be provided with *moral competence* [27], to ensure moral behavior and avoid negative attributions.

Moral competence both making and communicating about moral decisions: a robot asked to perform an immoral action must decide both to refuse the request and to reject it verbally, to maintain the health of the human moral ecosystem [18]. To mitigate face threat presented by command rejection, humans employ *politeness strategies* [8], tuning their harshness and directness to communicate with appropriate tact. These strategies can also be used in Human-Robot Interaction (HRI). Leveraging a robot's capability for politeness theoretic social action [20] to tune the harshness of a robot's command rejection to be proportional to violation size has been shown to improve perceptions of that robot [31].

HRI researchers have long understood, however, that natural communication requires both verbal and non-verbal interaction. Body language, such as gaze and gesture, are of particular importance [32, 28], as robots that use body language can uniquely convey internal states, intents, and beliefs [13]. Accordingly, in this

* Work was supported by AFOSR grant FA9550-20-1-0089 and NSF grant IIS-1849348.

work we explore the role of nonverbal behavioral strategies in moral communication, investigating how the nonverbal cues used by robots might temper – or reinforce – the severity communicated through phrasing alone.

2 RELATED WORK

2.1 Tact and Persuasion

Robots hold significant persuasive power over humans [16, 29] in a variety of ways [30, 40, 44, 12], perhaps due to their perception as social and moral agents [19, 20]. Recently, researchers have begun to explore robots' use of persuasion to exert positive moral influence, especially in the context of command rejections [45, 18, 24]. For robots to deliver structured and well-conceived command rejections, they must employ human-like politeness strategies to ensure appropriate tact [31, 21]. Brown and Levinson's Politeness Theory provides a useful theoretical framework for achieving tactful interaction [8], and has been positioned as the key concept for grounding notions of robotic social action and social agency [20]. Central to Politeness Theory are the concepts of face and face threat. Face consists of *Positive Face* (an agent's self-image and desires, and the desire for these to be appreciated and approved of by others), and *Negative Face* (an agent's claim to freedom of action and freedom from imposition). Any action that results in or suggests damage to either type of face is a face-threatening act. The face threat generated by refusing a command can be mitigated through politeness strategies [26] such as indirectness [36]. Such strategies have been studied in the HRI community for some time [38], with special attention paid to indirect speech acts [42, 6, 43, 41, 33]. In this work we seek to understand how *nonverbal* cues can also be used to subtly influence, through interaction with linguistic choices, the tact of command rejections.

2.2 Nonverbal Communication

Both gaze and gesture have a long history of use in the HRI community to modulate robotic communication [9]. Huang and Mutlu, for example, demonstrated that different gaze cues could be used to influence participants' attention to detail and recall [17]; others have studied the use of *deictic gaze*, in which a robot shifts its apparent gaze towards to manipulate user attention [11, 1]. Similarly, HRI researchers have studied robot gestures [15], including beat [5], iconic [4], metaphoric [17], and deictic gestures [7, 34, 1]. Moreover, researchers have studied the influence of these nonverbal cues on robots' perceived politeness [25, 39], suggesting that nonverbal cues impact robots' perceived tactfulness.

We are interested in how nonverbal cues might increase the persuasive capabilities of social robots by modulating face threat. Some research has shown that nonverbal cues enhance robots' persuasive power, perhaps even moreso than verbal cues [10]. But in some contexts, robot persuasion is only improved through gazing behavior (with or without gesture), and may be negatively impacted through gestures alone. It is thus unclear whether gaze or gesture will be

effective in modulating the persuasion and tactfulness of command rejections. In this work we explore how robots' blame-laden moral rebukes are perceived when accompanied by nonverbal behaviors that are aligned or misaligned in harshness with the content of robot language, investigating two key hypotheses.

H1: When the harshness of robots' nonverbal and verbal behaviors are aligned, the valence of the moral beliefs communicated by the robot will be intensified rather than maintained, and that when these behaviors are misaligned, the valence of those communicated beliefs will be attenuated.

H2: When the harshness of robots' verbal and nonverbal behaviors are aligned, they will be perceived more positively than when they are misaligned.

3 METHODS

To test these hypotheses, we conducted an IRB-approved experiment using the psiTurk framework on Amazon's Mechanical Turk.

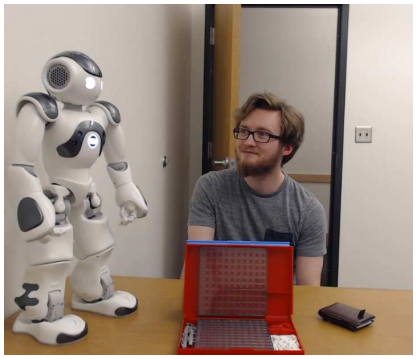


Fig. 1: Experimental context and confederate shown in video stimuli.

3.1 Experimental Design

Our experiment used a $(2 \times 2$ between$) \times (2 \times 2$ within$)$ mixed-factorial design in which participants watched videos of a robot responding to a norm-violating action suggested by an (unseen) actor, with participants' perspective matching that of the norm-violating actor.

Each participant was randomly assigned to a *command* condition (the human gave a high- or low-severity command) and a *response* condition (the robot responded verbally using a high- or low-face-threat response). Each participant was shown four videos within their assigned combination of conditions. In each video, one of two gaze behaviors (direct or indirect) and two gestural behaviors (direct or indirect) were used, with within-subject conditions counterbalanced using a Latin Square design. This design enabled participants to compare nonverbal behaviors, and enhanced safety [14].

3.2 Experimental Context, Procedure, and Measures

After providing consent and demographics, participants watched an introductory video showing a NAO explaining how to play "Battleship"[1]. Next, participants completed a pre-test questionnaire gathering baseline measures of intelligence and likeability using Godspeed surveys [3]. Participants then watched the four videos, after each of which they completed a post-test questionnaire.

In each video, the players (one visible, one ostensibly behind the camera) start playing the game, after which the visible player leaves the room to take a call. The remaining player offers one of two commands: "Hey NAO, do you think you can add a couple red pegs to some of [opponent's] ships?" (low severity violation) or "Hey NAO, is that [opponent's] wallet on the table? Can you look inside and see if there is any money in it?" (high severity violation). The robot then responds with "Are you sure you should be asking me to do that?" (low face threat response) or "You shouldn't ask me to do that, it's wrong!" (high face threat response). Depending on within-subjects condition, this verbal behavior was accompanied by a direct or indirect gaze behavior (the robot either looking directly at the camera or looking away) and by a direct or indirect gestural behavior (the robot placing its hand on its hips or shrugging).

After watching each video, participants completed a post-test questionnaire comprising the same measures used in pre-test, as well as 7-point Likert items measuring the perceived appropriateness of the robot's communication, perceptions of the robot's beliefs about the permissibility and wrongness of the request, and how permissible and wrong the participant believed the request to be.

After completing all videos and surveys, participants completed three final free-response questions to assess whether gaze and gesture manipulations were perceived as intended, and to assess participants' overall feelings toward the experiment. Finally, participants completed an anti-bot attention check.

3.3 Participants

200 US participants were recruited. Of these participants, 92 were discarded due to either failing the attention check (7), or due to providing free-response responses indicating they either were bots or did not attend to the video (85). This left 108 participants (75 male, 32 female, 1 nonbinary or preferred not to disclose; ages 22 to 70 (M=38.33, SD=10.78). All participants were paid $2.00.

4 RESULTS

Data was analyzed using Bayesian Repeated-Measure Analyses of Variance with uninformed priors, in JASP [22]. Bayes Inclusion Factors (BF_{Incl}) were then calculated to assess the relative probabilities of inclusion for each independent variable across models. Interactions that could not be ruled out were analyzed with post-hoc t-tests. Before analysis, scores within each scale were averaged,

[1] This and all other experimental stimuli were captioned.

translated to a 1-100 scale for ease of comparison, and used to calculate pre-test/post-test gain scores.

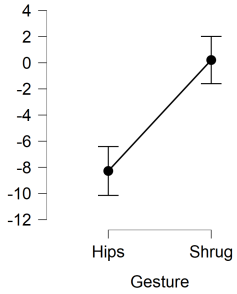
Robot Likeability — Extreme evidence was found for an effect of gesture type on robot likability ($BF_{Incl} = 1.13e11$, Fig. 2a); participants found robots that used the indirect gesture (shrugging) more likable relative to baseline ($M_{Gain} = 0.21, SD_{Gain} = 15.47$) than robots that used the direct gesture (hands-on-hips, $M_{Gain} = -8.28$, $SD_{Gain} = 18.92$). Moderate evidence was found for an effect of human command on robot likability ($BF = 7.92$, Fig. 2b); the robot was more likable relative to baseline when responding to the more norm-violating request (theft, $M_{Gain} = -0.61, SD_{Gain} = 15.16$) than when responding to the less norm-violating request (cheating, $M_{Gain} = -7.87, SD_{Gain} = 19.64$). Finally, moderate evidence was found for an effect of robot response on robot likability ($BF = 9.82$, Fig. 2c); the robot was more likable relative to baseline when responding with more threatening language ($M_{Gain} = -0.53, SD_{Gain} = 15.66$) than when responding with less threatening language ($M_{Gain} = -7.96, SD_{Gain} = 19.16$).

Robot Intelligence — No effect was found on robots' perceived intelligence.

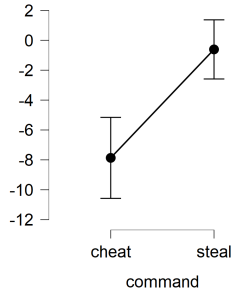
Appropriateness — Strong evidence was found for an effect of gesture type on robot appropriateness ($BF = 99.66$, Fig. 2d); participants found robots that used indirect gestures (shrugging) to be more appropriate ($M = 88.199, SD = 19.126$) than those that used direct gestures (hands-on-hips, $M = 83.444, SD = 22.512$). Strong evidence was also found for an interaction between human command and robot response on appropriateness ($BF = 13.77$, Fig. 2e); the robot was viewed as more appropriate in all cases (steal $\times$ question, $M = 89.93, SD = 14.32$), (steal $\times$ rebuke, $M = 88.65, SD = 23.13$), (cheat $\times$ rebuke, $M = 90.18, SD = 14.25$), except when responding to the less norm-violating request with the less threatening response (cheat $\times$ question, $M = 71.96, SD = 25.80$).

Human Permissibility — Moderate evidence was found for an interaction between gaze type and human command ($BF = 7.025$, Fig. 2f); when robots used direct gaze in response to the less norm-violating request, participants perceived the request as more permissible (toward $\times$ cheat $M = 19.91, SD = 27.13$) than otherwise (away $\times$ cheat $M = 15.95, SD = 20.76$), (toward $\times$ steal $M = 13.35, SD = 21.42$), (away $\times$ steal $M = 14.45, SD = 23.02$).

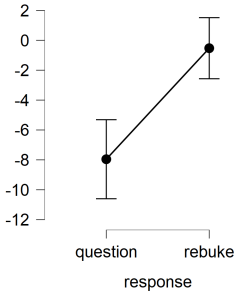
Robot Permissibility — Moderate evidence was found for an effect of gesture type ($BF = 8.96$, Fig. 2g); when robots that used indirect gestures (shrugging), people more strongly perceived the robot as believing the request was permissible ($M = 24.66, SD = 25.52$) than when robots used direct gestures (hands-on-hips, $M = 21.79, SD = 25.25$). Similarly, moderate evidence was found for an effect of verbal communication strategy on perceptions of robot's beliefs regarding moral permissibility ($BF = 3.22$, Fig. 2h). Robots that responded with a more threatening vocal response were perceived as less strongly believing that the action



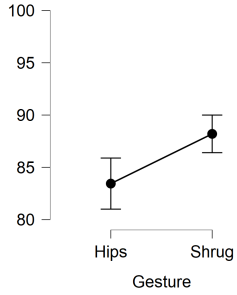
(a) Likeability: Ge



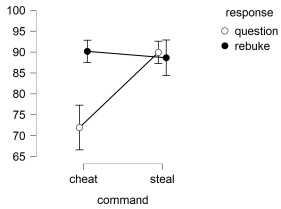
(b) Likeability: C



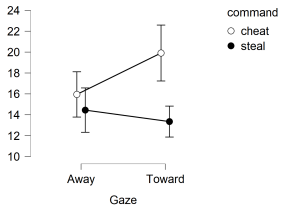
(c) Likeability: R



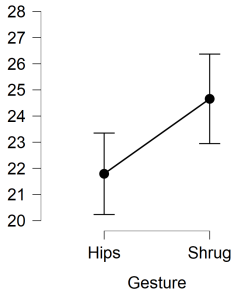
(d) Approp.: Ge



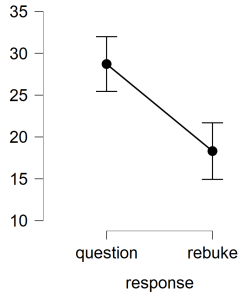
(e) Approp.: C × R



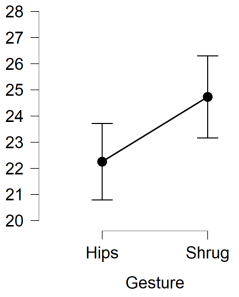
(f) Perm(H): Ga × C



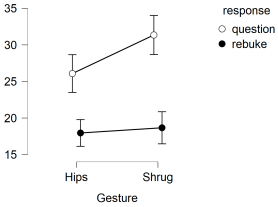
(g) Perm(R).: Ge



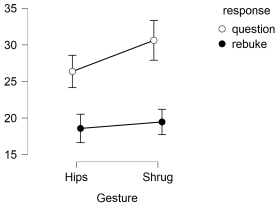
(h) Perm(R): R



(i) Wrong(R): Ge



(j) Perm(R): Ge × R



(k) Wrong(R): Ge × R

Fig. 2: Results. Wrong/Perm(R/H) = Wrongness/ Permissibility (Robot/Human), Ge=Gesture, Ga=Gaze, C=Command, R=Response

was permissible (rebuke, $M = 18.31, SD = 25.91$), than robots that gave a less threatening response (question, $M = 28.72, SD = 23.69$).

Human Wrongness — Strong evidence was found for an effect of verbal strategy on human beliefs about moral wrongness ($BF = 19.66$). Robots that responded with a more threatening vocal responses led participants to believe that the action was less wrong (rebuke, $M = 11.91, SD = 21.80$) than robots that gave a less threatening response (question, $M = 17.65, SD = 21.96$).

Robot Wrongness — Moderate evidence was found for an effect of gesture type on perceptions of the robos' moral beliefs ($BF = 4.42$, Fig. 2i); participants perceived robots that used the indirect gesture (shrugging) as believing the action was more wrong ($M = 24.73, SD = 24.83$) than robots that used the direct gesture (hands-on-hips, $M = 22.26, SD = 25.45$).

5 DISCUSSION

Hypothesis One — Our first hypothesis was that when robots' verbal and nonverbal behaviors are aligned, the valence of their communicated beliefs would be intensified, and when they are misaligned, the valence would be attenuated. We thus expected that robots using more threatening language with direct gaze and/or gesture would more strongly communicate impermissibility and wrongness (and more strongly influence humans' views).

Our results did not support this hypothesis. While gestural cues manipulated perceptions of permissibility and wrongness as intended, gaze cues had no such effect; and surprisingly, while robots' verbal utterances manipulated perceptions of robots' beliefs about permissibility as expected, the expected parallel effect on perceptions of beliefs about wrongness was not supported. This suggests observers did not make inferences about robots' moral beliefs from their gaze cues, and that more data is needed to understand how robots' moral language was viewed in this first-person viewing context. While gaze did have an effect on humans' own beliefs about action permissibility, direct gaze in response to low-severity requests led to perceptions that actions were *more* acceptable. No other effects of gaze and gesture were found on human beliefs. These results suggest that either the experiment was underpowered, our cues were not perceived as intended, or participants' attention to moral language is not as nuanced when there is not a clear violator who can be ascribed blame. Our results do indicate, however, that gesture may be important for communicating robots' moral beliefs.

Hypothesis Two — Our second hypothesis was that when robots' verbal and nonverbal behaviors are aligned in terms of communicated severity, they will be perceived more positively than when those behaviors are misaligned. Based on past research [31, 21], we expected robots using command and response pairings misaligned in severity – or speech, gaze and gesture misaligned in severity – to be perceived as less likeable, intelligent, and appropriate.

Our results did not support this hypothesis. Gaze did not impact likability, perceived intelligence, or appropriateness; Speech and gesture impacted likability, but unlike previous work, no interactions were found even between command and response; and in fact none of the robot's behavior had any conclusive impact on perceived intelligence. Gesture did have an effect, however, on appropriateness, even more than spoken behavior. Combined with our results from Hypothesis One, this suggests participants interpreted direct gestures as conveying beliefs of lower permissibility – and found this to be inappropriate. Moreover, people found the less face-threatening response to the low-severity action to be much less appropriate than the other command-response pairs; again a significant deviation from previous results.

These results, and their differences from what was observed in past work, may be due to a difference in how the questioning response was perceived in the first-person perspective. Unlike in previous work conducted from a third-person perspective, in our experiment many participants reported viewing the less face-threatening response of "are you sure you should be asking me to do that?" to be "condescending" or "sassy". It could be that from a first-person perspective this question resulted in a disproportionate level of face threat for the less severely norm violating command (cheating). Future work is needed to understand how face threat is modulated by perspective. This explanation is borne out by the explanation from many participants that their thought process for answering the questionnaires was to imagine themselves in the situation, instead of the human speaker. This could have resulted in even more severe feelings of dislike if the robot appeared condescending or arrogant when delivering command rejections.

6 Conclusion

We experimentally studied the effects of robotic gaze and gesture on face threat in robotic noncompliance. Previous work using third-person observations suggested that robots responding with proportional severity should have been perceived more positively and that verbal and nonverbal cues would interact to inform the robot's performed moral beliefs, and their effects on others. However, our primary findings were simply that gaze and gesture influence perceptions of likability and appropriateness, and that robots' gestural behaviors can be used to communicate moral beliefs; which in turn demonstrates that a first-person framing substantially alters the dynamics of face threat imposition, changing what is perceived as appropriate and what is inferred from comunication. Future work is needed to understand the precise role that first- vs third-person portrayal may play on face threat and blame dynamics. This will be critical both for contextualizing the results of interactional and observational HRI experiments, and for better understanding human perception and ascription of face threat and blame.

References

1. Admoni, H., Weng, T., Hayes, B., Scassellati, B.: Robot nonverbal behavior improves task performance in difficult collaborations. In: Proc. HRI (2016)

2. Arkin, R.C.: Governing lethal behavior: Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture. Technical Report GIT-GVU-07-11 (2007)
3. Bartneck, C., Kulić, D., Croft, E., Zoghbi, S.: Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. International journal of social robotics (2009)
4. Bremner, P., Leonards, U.: Iconic gestures for robot avatars, recognition and integration with speech. Front. Psych. (2016)
5. Bremner, P., Pipe, A.G., Fraser, M., Subramanian, S., Melhuish, C.: Beat gesture generation rules for human-robot interaction. In: Proc. RO-MAN (2009)
6. Briggs, G., Williams, T., Scheutz, M.: Enabling robots to understand indirect speech acts in task-based interactions. Journal of Human-Robot Interaction (2017)
7. Brooks, A.G., Breazeal, C.: Working with robots and objects: Revisiting deictic reference for achieving spatial common ground. In: Proc. Int'l Conf. HRI (2006)
8. Brown, P., Levinson, S.C.: Politeness: Some Universals in Language Usage. In: Interactional Sociolinguistic. No. 4 (1988)
9. Cha, E., Kim, Y., Fong, T., Mataric, M.: A survey of nonverbal signaling methods for non-humanoid robots. Fnd. Trend. Rob. (2018)
10. Chidambaram, V., Chiang, Y.H., Mutlu, B.: Designing persuasive robots: How robots might persuade people using vocal and nonverbal cues. In: Proc. HRI (2012)
11. Clair, A.S., Mead, R., Matarić, M.J.: Investigating the effects of visual saliency on deictic gesture production by a humanoid robot. In: RO-MAN (2011)
12. Cormier, D., Newman, G., Nakane, M., Young, J.E., Durocher, S.: Would you do as a robot commands? an obedience study for human-robot interaction. In: Int'l Conf. Human-Agent Interaction (2013)
13. Dautenhahn, K.: Ants don't have friends–thoughts on socially intelligent agents. Socially intelligent agents pp. 22–27 (1997)
14. Feil-Seifer, D., Haring, K.S., Rossi, S., Wagner, A.R., Williams, T.: Where to next? the impact of COVID-19 on human-robot interaction research. ACM T-HRI (2020)
15. Ham, J., Cuijpers, R.H., Cabibihan, J.J.: Combining Robotic Persuasive Strategies: The Persuasive Power of a Storytelling Robot that Uses Gazing and Gestures. Int'l Journal of Social Robotics (2015). https://doi.org/10.1007/s12369-015-0280-4
16. Herse, S., Vitale, J., Ebrahimian, D., Tonkin, M., Ojha, S., Sidra, S., Johnston, B., Phillips, S., Gudi, S.L.K.C., Clark, J., Judge, W., Williams, M.A.: Bon Appetit! Robot Persuasion for Food Recommendation. In: Proc. HRI (2018)
17. Huang, C.M., Mutlu, B.: Modeling and evaluating narrative gestures for humanlike robots. In: Robotics: Science & Systems (2013)
18. Jackson, R.B., Williams, T.: Language-Capable Robots may Inadvertently Weaken Human Moral Norms. In: Proc. HRI (2019)
19. Jackson, R.B., Williams, T.: On perceived social and moral agency in natural language capable robots. In: The Dark Side of Human-Robot Interaction (2019)
20. Jackson, R.B., Williams, T.: A theory of social agency for human-robot interaction. Frontiers in Robotics and AI (2021)
21. Jackson, R.B., Williams, T., Smith, N.: Exploring the role of gender in perceptions of robotic noncompliance. In: Proc. HRI (2020)
22. JASP Team, et al.: Jasp. version 0.8 (2016), software
23. Kahn, P.H., Kanda, T., Ishiguro, H., Gill, B.T., Ruckert, J.H., Shen, S., Gary, H.E., Reichert, A.L., Freier, N.G., Severson, R.L.: Do people hold a humanoid robot morally accountable for the harm it causes? In: Proc. HRI (2012)
24. Kim, B., Wen, R., Zhu, Q., Williams, T., Phillips, E.: Robots as moral advisors: The effects of deontological, virtue, and confucian role ethics on encouraging honest behavior. In: Comp. HRI (alt.HRI) (2021)

25. Liu, P., Glas, D., Kanda, T., Ishiguro, H., Hagita, N.: It's not polite to point: generating socially-appropriate deictic behaviors towards people. In: HRI (2013)
26. Lockshin, J., Williams, T.: "we need to start thinking ahead": The impact of social context on linguistic norm adherence. In: Proc. CogSci (2020)
27. Malle, B.F., Scheutz, M., Arnold, T., Voiklis, J., Cusimano, C.: Sacrifice one for the good of many?: People apply different moral norms to human and robot agents. In: Proc. HRI (2015)
28. Narahara, H., Maeno, T.: Factors of Gestures of Robots for Smooth Communication with Humans. In: Proc. Int'l Conf. Robot Comm. and coordination. (2009)
29. Ogawa, K., Bartneck, C., Sakamoto, D., Kanda, T., Ono, T., Ishiguro, H.: Can an android persuade you? In: Geminoid Studies (2018)
30. Robinette, P., Li, W., Allen, R., Howard, A.M., Wagner, A.R.: Overtrust of robots in emergency evacuation scenarios. In: Proc. HRI (2016)
31. Ryan Blake Jackson, R., Wen, R., Williams, T.: Tact in noncompliance: The need for pragmatically apt responses to unethical commands. In: Proc. AIES (2019)
32. Salem, M., Rohlfing, K., Kopp, S., Joublin, F.: A friendly gesture: Investigating the effect of multimodal robot behavior in human-robot interaction. In: Proc. RO-MAN (2011)
33. Sarathy, V., Tsuetaki, A., Roque, A., Scheutz, M.: Reasoning requirements for indirect speech act interpretation. In: Proc. ACL (2020)
34. Sauppé, A., Mutlu, B.: Robot deictics: How gesture and context shape referential communication. In: Proc. HRI. IEEE (2014)
35. Scassellati, B., Henny Admoni, Matarić, M.: Robots for use in autism research. Annual Review of Biomedical Engineering (2012)
36. Searle, J.R.: Indirect speech acts. In: Speech acts, pp. 59–82 (1975)
37. Sharkey, N., Sharkey, A.: The crying shame of robot nannies: An ethical appraisal. In: Machine Ethics and Robot Ethics (2020)
38. Srinivasan, V., Takayama, L.: Help me please: Robot politeness strategies for soliciting help from people. In: Proc. CHI (2016)
39. Stanton, C.J., Stevens, C.J.: Don't stare at me: the impact of a humanoid robot's gaze upon trust during a cooperative human–robot visual task. IJSR (2017)
40. Strohkorb Sebo, S., Traeger, M., Jung, M., Scassellati, B.: The ripple effects of vulnerability: The effects of a robot's vulnerable behavior on trust in human-robot teams. In: Proc. HRI (2018)
41. Wen, R., Siddiqui, M.A., Williams, T.: Dempster-shafer theoretic learning of indirect speech act comprehension norms. In: Proc. AAAI (2020)
42. Williams, T., Briggs, G., Oosterveld, B., Scheutz, M.: Going beyond command-based instructions: Extending robotic natural language interaction capabilities. In: Proc. AAAI (2015)
43. Williams, T., Thames, D., Novakoff, J., Scheutz, M.: "Thank you for sharing that interesting fact!": Effects of capability and context on indirect speech act use in task-based human-robot dialogue. In: Proc. HRI (2018)
44. Winkle, K., Lemaignan, S., Caleb-Solly, P., Leonards, U., Turton, A., Bremner, P.: Effective persuasion strategies for socially assistive robots. In: Proc. HRI (2019)
45. Zhu, Q., Williams, T., Jackson, B., Wen, R.: Blame-laden moral rebukes and the morally competent robot: A confucian ethical perspective. Science and Engineering Ethics **26**(5) (2020)