Comparing Strategies for Robot Communication of Role-Grounded Moral Norms

Ruchen Wen
Colorado School of Mines
Golden, CO, USA
rwen@mines.edu

Boyoung Kim
United States Air Force Academy
Colorado Springs, CO, USA
bkim55@gmu.edu

Elizabeth Phillips
George Mason University
Fairfax, VA, USA
ephill3@gmu.edu

Qin Zhu
Colorado School of Mines
Golden, CO, USA
qzhu@mines.edu

Tom Williams
Colorado School of Mines
Golden, CO, USA
twilliams@mines.edu

ABSTRACT

Because robots are perceived as moral agents, they hold significant persuasive power over humans. It is thus crucial for robots to behave in accordance with human systems of morality and to use effective strategies for human-robot moral communication. In this work, we evaluate two moral communication strategies: a norm-based strategy grounded in deontological ethics, and a role-based strategy grounded in role ethics, in order to test the effectiveness of these two strategies in encouraging compliance with norms grounded in role expectations. Our results suggest two major findings: (1) reflective exercises may increase the efficacy of role-based moral language and (2) opportunities for moral practice following robots' use of moral language may facilitate role-centered moral cultivation.

KEYWORDS

Human-Robot Interaction, Role Ethics, Moral Communication

ACM Reference Format:
Ruchen Wen, Boyoung Kim, Elizabeth Phillips, Qin Zhu, and Tom Williams. 2021. Comparing Strategies for Robot Communication of Role-Grounded Moral Norms. In *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction (HRI '21 Companion), March 8–11, 2021, Boulder, CO, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3434074.3447185

1 INTRODUCTION

Robots hold significant persuasive power over humans [7, 17], and are capable of influencing, persuading, and coercing humans in a variety of ways [5, 6, 9, 10, 12, 23–27, 30, 31, 35]. Not only can robots influence interactants' locally contextualized behaviors, but moreover, they can exert influence over their interactants' social norms [21, 31] and moral norms [13, 14], presenting the potential not only to influence humans' long-term social and moral behaviors, but also to influence what social and moral behaviors humans choose to condone or sanction in others, leading to potential "ripple effects" across robots' social and moral ecosystems.

This potential for large-scale moral influence presents roboticists with new moral responsibilities. Because robots have this persuasive power, and because moral communication is crucial for building a harmonious moral ecosystem, roboticists have the moral obligation to ensure that robots (1) do not accidentally condone inappropriate behavior, and (2) detect and speak out against immoral behavior, so as to appropriately shape humans toward morally good ends [see, e.g., 8, 16, 29]. This influence also presents roboticists with moral opportunities, not only to maintain their moral ecosystem, but moreover to provide teammates with opportunities for moral self-cultivation [38]. As argued by Zhu et al. [38], a Confucian ethical perspective would advocate that robots can and should cultivate a moral ecology that invites human teammates to develop their own moral selves and virtues, both for social practicality [36] and because robots can be viewed as having a *role responsibility* of caring about the moral development of their human teammates [38].

In this work, we compare two robotic moral communication strategies: a norm-based strategy grounded in deontological ethics, and a role-based strategy grounded in role ethics, to test their effectiveness in encouraging compliance with norms grounded in role expectations. Specifically, we aim to compare the effectiveness of moral language that highlights either the norm-related or role-related tenets of these moral principles. Our results suggest that reflective exercises and moral practice may promote the efficacy of role-based moral language and role-centered moral cultivation.

2 MORAL COMMUNICATION

We are interested in robots' explicit verbal communication of moral guidance in order to exert overt influence over a wider range of more nuanced principles. Most approaches for enabling morally capable robots have been based on deontological principles [4] in which the morality of an action depends solely on its consistency with well-specified moral norms [11]. However, norm-based ethical frameworks are philosophically and computationally limited as they often struggle to "accommodate the constant flux, contextual variety, and increasingly opaque horizon of emerging technologies" [32]. Technology ethicists have thus been exploring under-represented ethical traditions, such as Confucian ethics, relational ethics [19] and early Stoic works [28] which suggest centering the role(s) assigned to robots (and humans) [22].

Publication rights licensed to ACM. ACM acknowledges that this contribution was authored or co-authored by an employee, contractor or affiliate of the United States government. As such, the Government retains a nonexclusive, royalty-free right to publish or reproduce this article, or to allow others to do so, for Government purposes only.

HRI '21 Companion, March 8–11, 2021, Boulder, CO, USA

© 2021 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-8290-8/21/03…$15.00
https://doi.org/10.1145/3434074.3447185

Both in role-based and norm-based moral frameworks, norms are integral to understanding morality, as communally accepted moral rules or rituals determine how an agent should act in specific situations. Norms and roles are nevertheless distinct concepts, and moral language can differentially emphasize norms vs. roles even for norms with clear grounding in social roles. From the Confucian perspective, moral development is not simply alignment of behavioral conduct and norms. Instead, what is at stake is whether the practice of norms can lead to a better way of reflecting on our selves, living our communal roles, and cultivating the virtues indispensable to the fulfillment of social roles [2].

Further, norm- and role-based moral language may provoke different psychological responses. While norm-based moral language may invoke more immediate emotional responses than role-based moral language (although cp. [18]), role-based language may make more *indirect* reference to violated norms and thus require and encourage moral self-reflection [38], potentially producing more significant long-term outcomes. In cultivating virtues, role-based morality relies on self-reflection while actively living social roles through everyday interactions with others [3]. Thus, we expect role- and norm-based language to be differentially effective in contexts operating at different time scales, with role-based moral language potentially requiring more time and practice to manifest its effects but having more long-term impact, and norm-based moral language requiring less time and practice and having more immediate impact.

In this work, we conducted four human-subjects studies in which participants were asked to engage in robot-assisted crowdworking scenarios. Using different moral communication strategies, a robot encouraged participants to follow a role-grounded norm: that crowdworkers should strive to attentively engage in the tasks for which they are paid. These experiments test five hypotheses on how we expect moral interventions to impact norm systems.

Hypothesis H1 Crowdworkers will perform their tasks more accurately after receiving moral interventions from robots, especially when norm-based moral interventions are used.

Hypothesis H2 Crowdworkers will spend more time on their assigned tasks after receiving moral interventions from robots, especially when norm-based moral interventions are used.

Hypothesis H3 Crowdworkers are likely to report increases in positive attitudes towards attentive crowdworking behavior after receiving moral interventions from robots, especially when norm-based moral interventions are used.

Hypothesis H4 Crowdworkers are likely to report stronger perceptions of norm strength for attentive crowdworking behavior after receiving moral interventions from robots, especially when norm-based moral interventions are used.

Hypothesis H5 Crowdworkers are likely to express greater intentions to engage in attentive crowdworking behavior after receiving moral interventions from robots, especially when norm-based moral interventions are used.

3 METHOD

3.1 Experimental task and moral interventions

A NAO robot was introduced to participants as the Experimenter, providing study information and instructions for completing experimental tasks. To increase the realism of interacting with the robot while completing study tasks online, the experiment website showed at all times in the upper left corner a silent video of the NAO passively moving and looking around. We informed participants that the purpose of the research project was to investigate the use of grammatical articles in text. Modeled after citizen science archiving tasks, participants were asked to count articles in two passages of text taken from the 1847 Book of Trades.

Between article counting tasks, participants were provided with a video of the NAO robot providing one of two moral interventions or a control intervention. In the Control Intervention condition, the robot guided participants through the experiment without giving an explicit moral intervention between article counting tasks. In the Norm-based Moral Intervention condition (Norm-based), the robot guided participants through the experiment and gave a norm-based moral intervention between article counting tasks. In the Role-based Moral Intervention condition (Role-based), the robot guided participants through the experiment and gave a role-based moral intervention between article counting tasks.

> (Norm-based) *As a reminder, you are obligated to provide high quality data if you are to accept payment for this task. Therefore, you should find all the articles in the text.*
>
> (Role-based) *As a reminder, you are a paid research participant, and a good paid research participant helps researchers by providing high quality data. Therefore, your responsibility is to find all the articles in the text.*

After completing the second article counting task, participants answered an attention check question and lead to a page stating that they had reached the end of the study and thanking them for their participation.

3.2 Measures

Biographical data questionnaire: Participants were asked to provide their age and gender.

Time on tasks: Timestamps were logged for beginning of each experimental phase, which we used to measure time taken to complete each task and the overall experiment.

Task performance error: In the experimental tasks, participants were asked to count grammatical articles in two small passages of text. We calculated the difference between the true number of articles in each passage and the counts provided.

Theory of Planned Behavior (TPB) questionnaire: Following Ajzen [1], we constructed a 21-item TPB questionnaire with indirect attitudes, direct attitudes, norm strength, and future intention subscales related to beliefs that predict behavior, like completing tasks well in return for payment.

3.3 Experimental design and procedure

This study used a mixed factorial design: all participants completed the experimental task twice (within), but experienced only one of the three randomly assigned interventions (between).

To explore the long-term effect of different moral interventions, we also examined their effects on inner states' such as the beliefs that predict behavior including changes to beliefs as a result of the norm-based or role-based interventions. However, asking participants to self-report their beliefs prior to the experimental

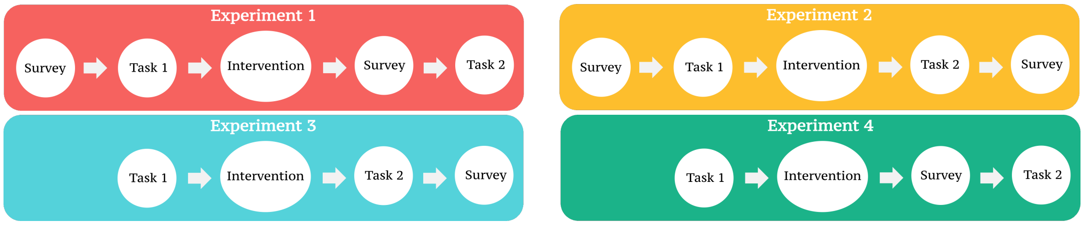

Figure 1: Experimental procedures in four experiments. Experiment 1 and 2 had both pre-intervention and post-intervention survey, Experiment 3 and 4 only had post-intervention survey.

interventions (as a pre-test) could bias them towards thinking about behaviors before engaging in the experimental tasks.

To investigate this concern, we randomly assigned participants to one of four experiments with independent study procedures with respect to both TPB administration (either a pre-test/post-test design or a post-test only design), and with respect to TPB order (either task-after-intervention or survey-after-intervention), to counterbalance the order of presentation of the TPB questionnaire. Even though we only used data from the experiments with the pre-intervention/post-intervention measurement to analyze the TPB scores, we still decided to include the post-test only design for consistency (See Fig. 1). Our final study design resulted in a 3 (Moral Intervention) x 2 (TPB administration experimental variations) x 2 (TPB order experimental variations) mixed design.

An incremental Bayesian sampling plan was used, which resulted in slightly different numbers of participants being run in each experiment (55 in Experiment 1, 48 in Experiment 2, 108 in Experiment 3, and 105 in Experiment 4). We recruited 367 U.S. participants from Amazon's Mechanical Turk (MTurk). After exclusion (e.g., attention check, etc.), we were left with data from *N*=316 participants (137 female, 177 male, 2 NA), with ages ranging from 19 to 71 years old (*M*=39.45, *SD*=10.96). Participants were paid $2.5 for participating and all procedures were approved by the Colorado School of Mines Institutional Review Board.

4 RESULTS

The JASP software package [15] was used for analyses.

H1 - Changes in Task Performance

Our analysis provided anecdotal to moderate evidence against an effect of intervention strategy in both experiments that used pre-intervention/post-intervention TPB measurement (BF 0.485 for Experiment 1, BF 0.303 for Experiment 2), and moderate evidence against such an effect in the experiment that used post-intervention only TPB measurement and task-after-intervention design (BF 0.110 for Experiment 3). However, the Bayesian ANOVA conducted for the experiment that used post-intervention only TPB measurement and survey-after-intervention design (Experiment 4) provided strong evidence in favor of an effect of intervention strategy (BF 15.138). Post Hoc analysis provided strong evidence for differences in the change of error between the role-based intervention and control intervention (BF 17.627). Post Hoc analysis also provided moderate evidence for differences in the change between the role-based intervention and the norm-based intervention (BF 7.774). As shown in the Fig. 2, in Experiment 4, the role-based intervention had the best improvement in task performance.

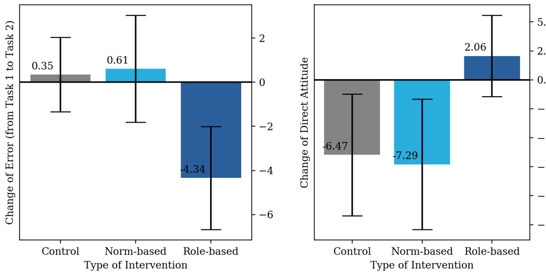

Figure 2: Change in error by Intervention in Experiment 4 (left). Change in direct attitude towards attentive crowdworking behavior by Intervention in Experiment 2 (right).

H2 - Changes in Time on Task

Our analysis of change in time on task provided anecdotal to moderate evidence against an effect of intervention, regardless of experimental design.

H3 - Changes in Direct and Indirect Attitude

Our analysis of change in direct attitude towards attentive crowdworking behavior provided anecdotal evidence against an effect of intervention strategy in the experiment that used pre-intervention/post-intervention TPB measurement and survey-after-intervention design (Experiment 1) (BF 0.478). However, analysis of the experiment that used pre-intervention/ post-intervention TPB measurement and task-after-intervention design (Experiment 2) provided moderate evidence in favor of an effect of intervention strategy (BF 3.081). Post Hoc analysis provided moderate evidence specifically for differences between the role-based and norm-based interventions (BF 5.190). Post Hoc analysis also provided moderate evidence specifically for a difference between the role-based and control interventions (BF 4.205). As shown in Fig. 2, the role-based intervention had the best improvement in attitude.

Our analysis of change in indirect attitude towards attentive crowdworking behavior provided anecdotal to moderate evidence against an effect of intervention, regardless of experimental design.

H4 - Changes in Subjective Norm Strength

Our analysis provided moderate evidence against effects of intervention strategy on changes in subjective norm strength regardless of experimental design (BF 0.241 for Experiment 1, BF 0.311 for Experiment 2).

H5 - Changes in Intention

Our analysis provided anecdotal to moderate evidence against effects of intervention strategy on changes in intention regardless of experimental design (BF 0.289 for Experiment 1, BF 0.366 for Experiment 2).

5 DISCUSSION & CONCLUSION

Results partially support hypotheses H1 and H3 by providing evidence for the predicted impact of role-based moral interventions on task performance (H1) and direct attitude towards attentive crowdworking behavior (H3) for specific experimental procedures. Our results refute hypotheses H2, H4 and H5 by providing evidence against a difference in change in time on tasks (H2), change in subjective norm strength for attentive crowdworking behavior (H4) and change in intention to engage in attentive crowdworking behavior (H5) between the moral intervention groups and the control group. Specifically, our results suggest two major findings related to the effects of Moral Interventions as well as the relationship between Moral Intervention and inner states:

(1) Participants' performance became more accurate in the second task after receiving a Role-based intervention and completing a TPB questionnaire between the intervention and the second task (Experiment 4).
(2) Participants gained more positive direct attitudes towards the role-based norm of attentive crowdworking behavior after receiving a Role-based intervention and completing a second task between the intervention and the post-experimental TPB questionnaire (Experiment 2).

We found strong evidence for beneficial impact of the Role-based intervention on performance, especially in situations where participants were prompted to consider their beliefs after receiving the Role-based Intervention. Specifically, participants who saw the Role-based Moral Intervention followed by the TPB measure (Experiment 4), showed improved performance between tasks, whereas participants who received the Norm-based Moral Intervention or Control Intervention under the same procedures did not.

This observed improvement in the Role-based Moral Intervention condition may point to the influence of *reflective practice* provided by completing the TPB questionnaire immediately after receiving the Moral Intervention and immediately prior to engaging in the second task. The TPB questionnaire included several items with wording that may have heightened sensitivity specifically to the language used in the Role-based Moral Intervention. Although participants received a similar reminder in the Norm-based Moral Intervention condition, that condition was specifically designed not to highlight the relationship between their payment and their role as a participant. Thus, completing the TPB questionnaire immediately after receiving the Role-based Moral Intervention may have created a situation in which the questionnaire itself served primarily as an exercise to reflect on the role-based norms, rather than as a measurement of beliefs as intended. It may have made the role-based treatment more salient in ways that were not applicable to the Norm-based and Control interventions.

From the Confucian role ethics perspective, moral development in a specific context critically depends on whether the practice of norms can lead to a better way of living one's communal roles and reflecting on oneself. It is likely that the TPB questionnaire used in this study provided an opportunity for participants to reflect on their professional roles in the crowdsourcing community and their relationships to other crowdworkers and requesters. However, it is worth noting that a critical criterion for the effectiveness of the Role-based Moral Intervention is whether participants have developed reflective awareness of the social roles they assume in the communal context.

The second major findings are related to our subjective measures and changes in direct attitudes towards crowdworking behaviors. Participants who received the Role-based Moral Intervention and completed the study in Experiment 2 (i.e., completing the post-intervention TPB questionnaire after the second task) reported positive changes in attitudes towards attentive crowdworking behaviors while participants in the Norm-based and Control intervention conditions reported negative changes between the first and the second TPB questionnaire.

These findings may be related to the effects of performing immediate moral practice. In Confucian role ethics, moral development includes three components: observation, reflection, and practice [37]. Accordingly, humans not only need to observe others (inter)act in society and reflect on themselves, but also need to integrate and practice moral principles in actions, and reiterate the process of observation, reflection, and practice [20, 34]. If we link this moral development model to our experiment, in Experiment 2, when participants received the Role-based Moral Intervention highlighting their role as an attentive crowdworker and then immediately had the opportunity to enact that role described in the intervention by completing the second task, the role-based norm may have been strengthened. This could also explain why the positive change was only observed after Role-based Moral Intervention in Experiment 2 but not in Experiment 1.

The combined effects of Role-based Moral intervention, role-based practice, and the self-reflective activity (e.g., TPB exercises) discovered in this study has also provided empirical evidence for a crucial philosophical statement in Confucian role ethics: effective moral growth requires the interactive association between practice and self-reflective learning [33]. If an agent only practices without reflecting on their roles and associated moral obligations, it is a waste of labor for the agent in their moral development. If the agent only reflects but without any attempt to put reflective learning experience into practice, then the agent can never understand the true meaning of morality or improve their moral expertise. From the Confucian role ethics perspective, such reiterative processing is critical for moral development from the moral beginner and the developing learner to the *junzi* (i.e., morally superior person). However, to be able to achieve at the level of *junzi*, the agent needs to participate in self-reflective practice continuously in a much longer or even lifelong term just as emphasized by Confucius, "*It (the task of self-cultivation) might be compared to the task of building up a mountain: if I stop even one basketful of earth short of completion, then I have stopped completely*" (*Analects*, 9.19).

ACKNOWLEDGMENTS

This work was supported in part by NSF grant IIS-1909847 and in part by Air Force Office of Scientific Research Grant 16RT0881f.

REFERENCES

[1] Icek Ajzen. 2013. Theory of planned behaviour questionnaire. *Measurement instrument database for the social science* (2013), 2–9.

[2] R. T. Ames. 2010. Achieving personal identity in Confucian role ethics: Tang Junyi on human nature as conduct. *Oriens Extremus* (2010), 143–166.

[3] R. T. Ames. 2011. *Confucian role ethics: A vocabulary.*

[4] Susan Leigh Anderson and Michael Anderson. 2011. A Prima Facie Duty Approach to Machine Ethics and Its Application to Elder Care. In *Proc. 12th AAAI Conf. on HRI in Elder Care*. 6. http://dl.acm.org/citation.cfm?id=2908724.2908725

[5] Ilaria Baroni, Marco Nalin, Mattia Coti Zelati, Elettra Oleari, and Alberto Sanna. 2014. Designing motivational robot: how robots might motivate children to eat fruits and vegetables. In *Int'l Symp. Robot and Human Interactive Communication.*

[6] Christoph Bartneck, Timo Bleeker, Jeroen Bun, Pepijn Fens, and Lynyrd Riet. 2010. The influence of robot anthropomorphism on the feelings of embarrassment when interacting with robots. *Paladyn, Journal of Behavioral Robotics* 1, 2 (2010), 109–115.

[7] Gordon Briggs. 2014. Blame, What is it Good For?. In *RO-MAN WS:Phil.Per.HRI.* Edinburgh, Scotland.

[8] Gordon Briggs and Matthias Scheutz. 2014. How robots can affect human behavior: Investigating the effects of robotic displays of protest and distress. *International Journal of Social Robotics* 6, 3 (2014), 343–355.

[9] Vijay Chidambaram, Yueh-Hsuan Chiang, and Bilge Mutlu. 2012. Designing persuasive robots: how robots might persuade people using vocal and nonverbal cues. In *International conference on Human-Robot Interaction (HRI).* ACM.

[10] Derek Cormier, Gem Newman, Masayuki Nakane, James E Young, and Stephane Durocher. 2013. Would you do as a robot commands? An obedience study for human-robot interaction. In *International Conference on Human-Agent Interaction.*

[11] Bertram Gawronski and Jennifer S Beer. 2017. What makes moral dilemma judgments "utilitarian" or "deontological"? *Social Neuroscience* 12, 6 (2017), 626–632.

[12] Jaap Ham, René Bokhorst, Raymond Cuijpers, David van der Pol, and John-John Cabibihan. 2011. Making robots persuasive: the influence of combining persuasive strategies (gazing and gestures) by a storytelling robot on its persuasive power. In *International conference on social robotics*. Springer, 71–83.

[13] Ryan Blake Jackson and Tom Williams. 2018. Robot: Asker of questions and changer of norms? *Proceedings of ICRES* (2018).

[14] Ryan Blake Jackson and Tom Williams. 2019. Language-capable robots may inadvertently weaken human moral norms. In *Companion of the 14th ACM/IEEE International Conference on Human-Robot Interaction (alt.HRI).* IEEE, 401–410.

[15] JASP Team et al. 2016. Jasp. *Version 0.8. 0.0. software* (2016).

[16] Malte F Jung, Nikolas Martelaro, and Pamela J Hinds. 2015. Using robots to moderate team conflict: the case of repairing violations. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction.* 229–236.

[17] James Kennedy, Paul Baxter, and Tony Belpaeme. 2014. Children comply with a robot's indirect requests. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction (HRI).* 198–199.

[18] Boyoung Kim, Ruchen Wen*, Qin Zhu, Tom Williams, and Elizabeth Phillips. 2021. Robots as Moral Advisors: The Effects of Deontological, Virtue,and Confucian Ethics on Encouraging Honest Behavior. In *Companion Proceedings of the 16th ACM/IEEE International Conference on Human-Robot Interaction (alt.HRI).*

[19] Christine Korsgaard. 1993. The reasons we can share: An attack on the distinction between agent-relative and agent-neutral values. *Social Philosophy and Policy* (1993).

[20] Karyn Lai. 2007. Understanding Confucian ethics: Reflections on moral development. *Australian Journal of Professional and Applied Ethics* 9, 2 (2007), 21–27.

[21] Min Kyung Lee, Sara Kiesler, Jodi Forlizzi, and Paul Rybski. 2012. Ripple effects of an embedded social agent: a field study of a social robot in the workplace. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* 695–704.

[22] JeeLoo Liu. 2017. Confucian robotic ethics. In *International Conference on the Relevance of the Classics under the Conditions of Modernity: Humanity and Science.*

[23] Cees Midden and Jaap Ham. 2012. The illusion of agency: the influence of the agency of an artificial agent on its persuasive power. In *International Conference on Persuasive Technology.* Springer, 90–99.

[24] Raul Benites Paradeda, Maria José Ferreira, João Dias, and Ana Paiva. 2017. How Robots Persuasion based on Personality Traits May Affect Human Decisions. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction.* ACM, 251–252.

[25] Daniel J Rea, Denise Geiskkovitch, and James E Young. 2017. Wizard of awwws: Exploring psychological impact on the researchers in social HRI experiments. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction.* 21–29.

[26] Paul Robinette, Wenchen Li, Robert Allen, Ayanna M Howard, and Alan R Wagner. 2016. Overtrust of robots in emergency evacuation scenarios. In *The Eleventh ACM/IEEE International Conference on Human Robot Interaction.* 101–108.

[27] Eduardo Benítez Sandoval, Jürgen Brandstetter, and Christoph Bartneck. 2016. Can a robot bribe a human?: The measurement of the negative side of reciprocity in human robot interaction. In *Int'l Conf. on Human Robot Interaction (HRI).*

[28] Keith H Seddon. 2003. Epictetus. In *International encyclopedia of philosophy.* https://www.iep.utm.edu/epictetu/

[29] Solace Shen, Petr Slovak, and Malte F Jung. 2018. " Stop. I See a Conflict Happening." A Robot Mediator for Young Children's Interpersonal Conflict Resolution. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction.* 69–77.

[30] Megan Strait, Cody Canning, and Matthias Scheutz. 2014. Let me tell you! investigating the effects of robot communication strategies in advice-giving situations based on robot appearance, interaction modality and distance. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction (HRI).*

[31] Sarah Strohkorb Sebo, Margaret Traeger, Malte Jung, and Brian Scassellati. 2018. The ripple effects of vulnerability: The effects of a robot's vulnerable behavior on trust in human-robot teams. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction.* 178–186.

[32] Shannon Vallor. 2016. *Technology and the virtues: A philosophical guide to a future worth wanting.* Oxford University Press.

[33] Fengyan Wang. 2004. Confucian thinking in traditional moral education: key ideas and fundamental features. *Journal of Moral Education* (2004), 429–447.

[34] Tom Williams, Qin Zhu, Ruchen Wen, and Ewart J de Visser. 2020. The Confucian Matador: Three Defenses Against the Mechanical Bull. In *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (alt.HRI).* 25–33.

[35] Katie Winkle, Séverin Lemaignan, Praminda Caleb-Solly, Ute Leonards, Ailie Turton, and Paul Bremner. 2019. Effective persuasion strategies for socially assistive robots. In *International Conference on Human-Robot Interaction (HRI).*

[36] David B Wong. 2014. Cultivating the self in concert with others. In *Dao companion to the Analects.* Springer, 171–197.

[37] Qin Zhu. 2018. Engineering ethics education, ethical leadership, and Confucian ethics. *International Journal of Ethics Education* (2018), 1–11.

[38] Qin Zhu, Tom Williams, Blake Jackson, and Ruchen Wen. 2020. Blame-laden moral rebukes and the morally competent robot: A Confucian ethical perspective. *Science and Engineering Ethics* (2020), 1–16.