Top-Down Effects on Anthropomorphism of a Robot

Hailey Scherer (Hailey.A.Scherer.20@Dartmouth.Edu)
Program in Cognitive Science, Undergraduate, 002 Carpenter Hall
Hanover, NH 03755 USA

Jonathan Phillips (Jonathan.S.Phillips@Dartmouth.Edu)
Program in Cognitive Science, 002 Carpenter Hall
Hanover, NH 03755 USA

Abstract

Anthropomorphism, or the attribution of human mental states and characteristics to non-human entities, has been widely demonstrated to be cued automatically by certain bottom-up appearance and behavior features in machines. The potential for top-down effects to influence anthropomorphism remains underexplored—even as most people's exposure to robots prominently features linguistic descriptions, e.g. in common discourse, public media, and product advertising. The results of this online experiment suggest that top-down linguistic cues, including personal pronouns and verbs that imply or reference humanlike mental states, increase anthropomorphism of a robot—and that these top-down cues may be as important of an influence as bottom-up cues. Moreover, these results suggest that this increased anthropomorphism is associated with increased unwarranted expectations of the robot's capabilities and increased moral regard for the robot. As robots and other machines become more integrated into human society, it is more important to understand the extent to which top-down influences matter for our thought, talk, and treatment of robots.

Keywords: anthropomorphism; human-robot interaction; top-down framing; metaphors; agency; experience

Introduction

As robots become more integrated into our everyday lives, it becomes increasingly critical to understand our remarkable tendency to anthropomorphize (attribute human mental states and characteristics to) them. Designers' inclusion of certain features in their robots' appearance and functionality may be more likely to elicit these attributions from users. Moreover, communication of these conceptualizations of the robot, e.g. calling the robot a "he" instead of "it" or a "companion" instead of a "tool" in marketing or common discourse, may also affect users' tendency to attribute human capacities to the robot. Anthropomorphism might also have important implications for how we regard and treat these machines in our social and ethical frameworks: anthropomorphism has been associated with overestimations of robot capabilities (theoretically; e.g. Duffy, 2003) increased trust (e.g. Waytz, Heafner, & Epley, 2014), and moral care (e.g. Nijssen, Müller, van Baaren, & Paulus, 2019). To date, most research on anthropomorphism has focused on the interface-level appearance and behavioral cues that may drive anthropomorphism in a 'bottom-up' way. By contrast, little work has explored 'top-down' effects on anthropomorphism or compared these two kinds of influence. Critically, however, most of the general public's exposure to robots today occurs through common discourse, public media, and product marketing, and, in these cases, language describing the robot is typically paired with images or videos of the robot alone or interacting with other humans. Thus, it is important to examine the relative influence of bottom-up and top-down cues to anthropomorphism, and how these cues interact, in a similar context. We take up these underexplored questions.

Background

Anthropomorphism has been shown to be cued automatically by particular appearance and behavior features: these include *facial features* (especially eyes) (e.g. Gray & Wegner, 2012; Novikova & Watts, 2015), *body morphology features* like arms (e.g. Haring, Watanabe, & Mougenot, 2013), *natural language processing and generation* (Mitchell & Xu, 2015; O'Neal, 2018), *quality of movement* and *apparent autonomy* (e.g. Heider & Simmel, 1944), and *contingent interaction* (Johnson, 2003). Behavior cues alone seem to cause anthropomorphic responses, but the presence of both behavior and appearance cues seems to cause the strongest anthropomorphic response (Johnson, 2003).

Anthropomorphism is associated with the attribution of *experience* and *agency* (Gray, Gray, & Wegner, 2007). Experience capacities include emotional and bodily states, e.g. fear, joy, pain, and consciousness. Agency capacities include cognitive abilities such as thought, emotion recognition, communication, planning, and moral reasoning. There is some evidence that the *experience* dimension might be intuitively viewed both as more essential to humanness and as more lacking in machines—but also might be the dimension we are more likely to perceive when we anthropomorphize (Gray et al. 2007; Gray & Wegner, 2012). That said, people have been observed to attribute both humanlike agency and experience to non-human entities (e.g. Johnson, 2003; Thellman, Silvervarg, & Ziemke, 2017).

The downstream consequences of how we understand robots may obtain at the design level. Conceptualizing a streaming service as a bookstore, a library, or a television network, for example, has resulted in markedly different products, such as iTunes, Netflix, or Spotify, respectively (Richards & Smart, 2016). In a similar way, conceptualizing a robot as humanlike or as fulfilling a human role—such as *employee* or *companion*, rather than *appliance* or *advanced tool*—will shape its designed appearance and behavioral repertoire. In turn, these designed features may function as bottom-up anthropomorphic cues and will constrain the set of possible interactions users will expect to have (and can have) with the robot. Anecdotally, robots are already assigned

societal roles typically occupied by humans, e.g. *confidant* (Schulevitz, 2018), *citizen* (Stone 2017), or *companion* (Darling, 2017; Anki, 2019). Personal pronouns and verbs that imply or reference mental states like "thinking" or "feeling" are also already used in reference to robots (Forlizzi, 2007). Indeed, there is already explicit normative debate regarding the extent to which we should use such language to describe machines—it is thought that such metaphoric language may matter for our thought, talk, and treatment of them on personal, ethical, and legal levels (e.g. Shellenbarger, 2019; Bryson, 2009; Darling, 2017).

Despite the clear theoretical importance of how people conceptualize robots, there remains much work to be done to understand the factors that influence this process. Previous work has shown evidence that verbal descriptions can be used to prime participants' judgements and conception of robots as social agents or members of certain groups, including nationalities, etc., through stereotype activation (e.g. Eyssel & Kuchenbrandt, 2012); in addition, framing the zoomorphic robot Paro as a companion or pet that sleeps and feels pain has been shown to influence human behavior (Wada, Kouzuki, & Inoue, 2012). That said, surprisingly little research has investigated whether and how top-down linguistic cues affect anthropomorphism specifically (Darling, Nandy, & Breazeal, 2015; Matthews, Lin, Panganiban, & Long, 2019; Nijssen et al. 2019; Onnasch & Roesler, 2019; Wallkötter, Stower, Kappas, & Castellano, 2020; Waytz et al. 2014). Though the existing body of work points to the potential importance of the question, prior research has not allowed for the direct comparison of top-down and bottom-up influences on anthropomorphism. Moreover, no prior research has investigated these factors' interactive impact on anthropomorphism in the context most typical of how the general population is currently exposed to robots, namely videos of the robot by itself or interacting with another human accompanied by some linguistic description. To address these gaps, the present study simultaneously manipulated both top-down linguistic descriptions of a robot and its rich bottom-up cues with videos of the robot, including videos of the robot interacting with another human. Thus, the present study allows us to (i) better understand the top-down influence of language on anthropomorphism and its associated implications, including capability estimations and moral regard, (ii) compare the influence of top-down vs. bottom-up anthropomorphic cues, and (iii) investigate how these two distinct sources of information about humanlike agency and experience interact with one another.

Methods

Participants

Amazon Mechanical Turk was used to recruit 658 participants currently in the United States. Of these, 61 did not complete the study and were excluded based on this incomplete status. Other participants were excluded based on failure to successfully complete any of 2 comprehension checks or failure to complete open response questions with relevant and reasonably coherent responses. After these exclusions, 533 participants were left for inclusion in analysis. These participants ranged in age from 18-69 years old (M = 36.84); 62.9% identified as male.

Procedure & Materials

After giving informed consent and completing a basic demographics questionnaire, participants were randomly assigned to one of three different Description conditions (Anthropomorphic, Mechanistic, or Neutral) and one of four Video conditions (No Video, Robot Solo, Mechanistic Interaction, Social Interaction) in a between-subjects design. Participants were asked to read the respective description, and then watch (if applicable) the respective video. Then, only for participants in the video conditions with a video (all except No Video), they wrote a short response describing what they saw in the video. All participants then completed (in order): an expected capabilities questionnaire, an anthropomorphism questionnaire, a scenarios questionnaire, and a technology familiarity questionnaire.

Descriptions There were three possible descriptions a participant could have received. The ***Neutral*** description consisted of a short sentence describing the robot's height and width, which was also incorporated into the other two description types.[1] The ***Mechanistic*** description portrayed the robot in terms of its technical specifications as listed in the Anki Vector promotional material (Anki, 2019), e.g. "a powerful four-microphone array." Possible uses were also mentioned (i.e. "transport small objects") and only passive verbs were used (e.g. "can be used to") except for the information about height and width, which used a descriptive verb (i.e. "has a height of…"). Only impersonal pronouns were used ("it"). The ***Anthropomorphic*** description portrayed the robot using anthropomorphic language such as the robot's "name" ("Vector"), personal pronouns ("he"), verbs that imply mental states ("talk with you") or directly refer to mental states ("loves to") and humanlike personality descriptors (e.g. "smart"; "curious"; "playful"). The possible uses mentioned in the Mechanistic description were here phrased as abilities and motivated with a personality ascription (i.e. "He loves to help out: he can..."). See the supplemental materials (DOI below) for full texts.

Videos There were four Video Type conditions, including a no video condition. All videos of an Anki Vector robot and (in two conditions) a researcher were 2:08 minutes.

The Anki Vector robot is an autonomous and artificially intelligent 3.9" x 2.4" x 2.7" robot. It has two treads and four wheels, a "head" that can move or nod up and down, a display

[1] As the baseline Description Type condition, this neutral description was used instead of (for example) no description to hedge against the possibility that the participants who were in this condition and did not watch a video would be picturing a robot of more human size and dimensions, which might have an effect on their responses in the subsequent questionnaires.

screen featuring eyes that appear to blink and emote, and an actuator similar to "arms." It was selected for use in the videos due to its possession of established anthropomorphic bottom-up features, including eyes, quality of movement, natural language processing capability, and behavior that is contingent the environment and human user.

In the two interaction videos (mechanistic and social), the same researcher interacted with the robot. For all videos, participants were instructed to watch carefully and told they would later be asked to describe what happens in the video. The ***Robot Solo Video*** featured the Anki Vector robot autonomously moving around a desk and making sounds, with no humans present, at times interacting with various items such as a plastic folder. The ***Mechanistic Interaction Video*** featured a human attending to an off-screen work task, with the robot on the desk next to her. The human interacted with the robot mechanistically: for example, pushing the button on the top of the robot to activate it, asking the robot to move a cube or about the distance between two cities. In the ***Social Interaction Video***, the human seemed to treat the Anki Vector more like a social companion. The human interacted with the robot socially: for example, greeting the robot by its name, playing music for it to dance to, or giving it a "fist bump". The robot recognized the human and called her by name. See the supplement for video links and files.

Measures

Linguistic Analysis All participants except those in the No Video condition received an open response question that asked them to "Please briefly describe what you saw in the video in your own words." Responses were required to be above a minimum of 75 characters. Linguistic analysis of these responses was theoretically informed by Johnson (2003), Heider and Simmel (1994), and Fussell, Kiesler, Setlock, and Yew (2008). Responses were manually coded for the number of impersonal pronouns ("it") and personal pronouns ("he" or "they") made *in reference to the robot*, the number of verbs implying humanlike mental states and the number of verbs directly referencing a humanlike mental state used *of which the robot was the explicit grammatical subject*, and the number of humanlike descriptors (adjectives, adverbs) used *in reference to the robot*. The personal pronouns, verbs implying or referencing humanlike mental states, and humanlike descriptors were summed into a single score labeled "Linguistic Markers of Anthropomorphism" (LMA), a dependent variable for analysis.

Expected Capabilities Questionnaire. To test the hypothesis that anthropomorphism is associated with overestimations of robot capabilities (Duffy, 2003; Scheutz, 2012), participants were asked to estimate the probability that the robot would have certain capabilities, none of which were warranted to expect based on the available information. Eight mechanistic (e.g. "Scan my fingerprints") and eight anthropomorphic (e.g. "Understand my emotions") capabilities were presented. Item order was randomized.

Anthropomorphism Questionnaire Three scales were included. The *Metaphors Scale* asked participants to estimate the probability that they would use a given metaphorical term to describe the robot, e.g. "Companion" and "Tool." The *Agency and Experience Scale*, informed by Gray et al. (2007), asked participants to indicate their agreement with statements that attributed a given capacity to a robot, e.g. "This robot is capable of thinking." The *Pairwise Comparison Scale*, adapted and abbreviated from Bartneck et al. (2009), asked participants to make disjunctive judgements on a sliding scale about which of two descriptive terms most closely matched their impression of the robot, e.g. "Machinelike/Humanlike."

Trust and Moral Care Scenarios Participants also responded to questions following hypothetical scenarios designed to test their trust of or moral concern for the robot. The moral concern items posed moral dilemma-style scenarios asking the participant to indicate on a sliding scale the likelihood of their taking one of two actions to promote either the safety of a human or that of the robot. Choosing the latter was counted as an election to save the robot and taken to show moral concern for the robot.

Results

Factor Analysis

A factor analysis was conducted on the 21 items from the Anthropomorphism questionnaire. A principal component analysis with Varimax orthogonal rotation yielded three factors explaining a total of 68.55% of the variance in item responses: *Anthropomorphism* (explaining 48.25% of variance), *Advanced Tool Perceptions* (explaining 10.96% of variance), and *Advanced Toy Attributions* (explaining 9.34% of variance) (see supplement for more information). Loading most highly in the Anthropomorphism factor were all four attributed experience capacities (feeling affection, feeling pleasure, experiencing emotions, feeling pain), all four attributed agency capacities (understanding others' emotions, thinking, making decisions; and communicating, though the latter was barely significant), all anthropomorphic metaphors (e.g. partner), and all three items from the anthropomorphism pairwise comparison scale. Three mechanistic metaphors (e.g. tool) and the *communicate* and *making decisions* agency capacities loaded most highly in the *Advanced Tool Perceptions* factor.

The Effect of Description and Video

Anthropomorphism. A two-way between-subjects 3x4 analysis of variance (ANOVA) tested the effect of video and description on the extracted factor Anthropomorphism. This test yielded no significant interaction, but did reveal a significant effect of description, $F(2,521) = 5.15$, $p < .01$ (Figure 1). Post-hoc tests (Tukey's HSD) revealed that the anthropomorphic description caused significantly more anthropomorphism compared to the neutral, $p < .005$. The mechanistic and neutral description conditions did not

significantly differ. Thus, across video conditions featuring different bottom-up cues, the top-down anthropomorphic descriptions significantly increased anthropomorphism.

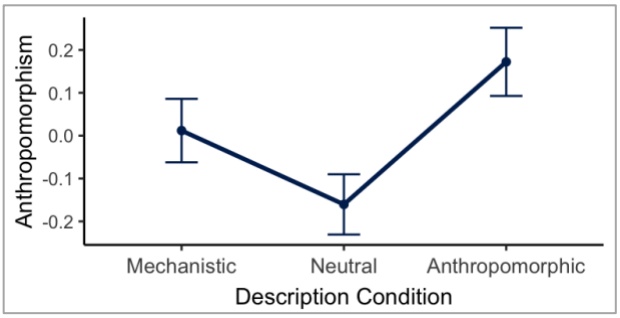

Figure 1: The effect of description on the extracted factor Anthropomorphism.

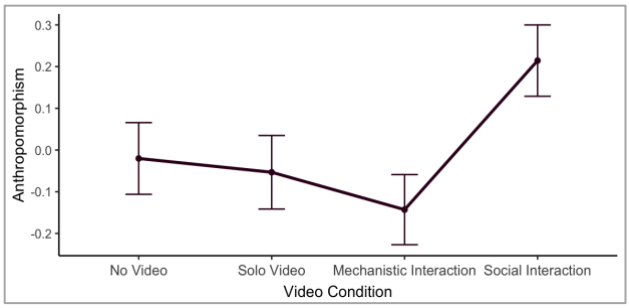

Figure 2: The effect of video on Anthropomorphism.

There was also an effect of video (bottom-up cues), $F(3,521) = 3.08$, $p < .05$, on Anthropomorphism (Figure 2). A one-way *t*-test also revealed that the social interaction video caused participants to explicitly anthropomorphize reliably more than the mean, $t(130) = 2.55$, $p < .05$. None of the other video conditions differed reliably from the mean.

Remarkably, especially in the face of the established research showing the importance of the effect of bottom-up cues on anthropomorphism, these results show that while these bottom-up cues do have an effect, top-down cues can have at least as large of an impact on anthropomorphism.

Agency and Experience. We also separately examined the effect of our manipulations on perceptions of *experience* and *agency*, allowing us to compare our results with perceptions of adult humans and a robot from Gray et al. (2007). For the attribution of experience, a two-way ANOVA yielded *only an effect of top-down cues* (description), $F(2, 521) = 5.34$, $p < .01$. Moreover, for the attribution of agency, a two-way ANOVA also yielded *only an effect of top-down cues*, $F(3, 521) = 4.44$, $p < .05$.

Our manipulations of bottom-up *and especially top-down* cues influenced perceptions of agency and experience. Indeed, as a result of our manipulations, participants conceived of the robot increasingly similarly to the way we conceive of human adults, as estimated by the Gray et al. (2007) study (Figure 3).

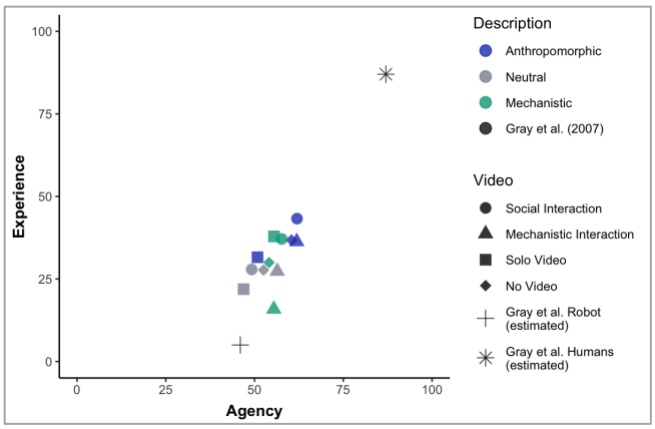

Figure 3: Condition-wise comparison of agency and experience scores compared to data from Gray et al. (2007).

Linguistic Analysis. Interrater reliability analysis measuring absolute agreement suggested high initial agreement between two independent raters, with intraclass correlation coefficient values ranging from .72 (95% CI [.66, .77]) to .96 (95% CI [.95, .96]) (see supplement for more information). Disagreements were resolved through consensus.

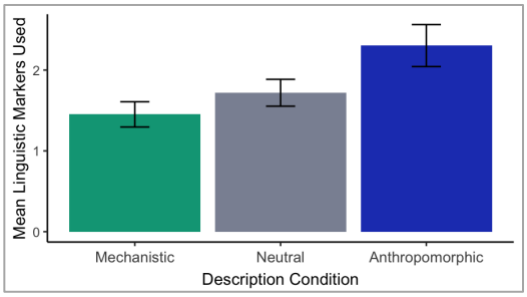

Figure 4: The effect of description on the use of LMA.

A two-way ANOVA testing the effect of description and video on linguistic markers of anthropomorphism (LMA) revealed only a significant effect of description, $F(2, 374) = 5.05$, $p < .01$ (Figure 4), and a significant effect of video, $F(2, 369) = 8.20$, $p < .001$. Post-hoc tests (Tukey's HSD) revealed that the anthropomorphic description caused the use of significantly more LMA compared to the mechanistic description, $p < .01$, and numerically more than the neutral description, $p = .07$. Post-hoc tests also showed the social interaction video caused the use of more LMA compared to both other groups, $p < .05$, and that the mechanistic interaction video caused the use of fewer markers compared to both other groups, $p < .05$.

Expected Capabilities. A two-way ANOVA yielded a significant effect of description, $F(2, 521) = 3.08$, $p < .05$, on participants' probability estimates that the robot would have certain ***anthropomorphic*** capabilities (Figure 5A). A two-way ANOVA also yielded a significant effect of description, $F(2, 521) = 3.20$, $p < .05$, on participants' probability estimates that the robot would have certain ***mechanistic***

capabilities (Figure 5B). Notably, the anthropomorphic description increased expectations of anthropomorphic capabilities compared to the other two conditions, while, likewise, the mechanistic description increased expectations of mechanistic capabilities compared to the other two conditions, suggesting the importance of top-down influences on expectations of corresponding capabilities.

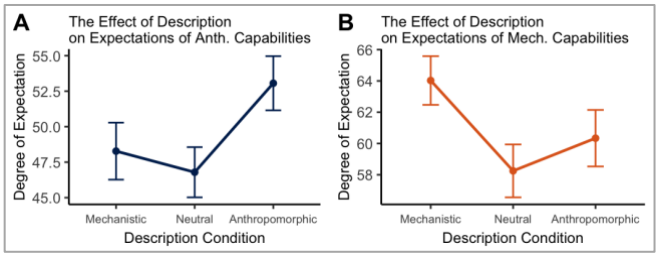

Figure 5: The effect of description on participants' expectations of robot capabilities.

There were also significant effects of video on expectations of anthropomorphic and mechanistic capabilities, $F(3, 521) = 5.00$, $p < .01$ and $F(3, 521) = 8.62$, $p < .001$, respectively, by which participants who watched the Robot Solo video had unusually low expectations overall and those who did not watch a video or watched the mechanistic interaction video had higher expectations of mechanistic capabilities.

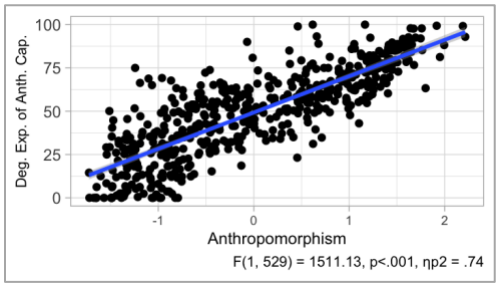

Figure 6: Anthropomorphism vs. Expectations of Anthropomorphic Capabilities.

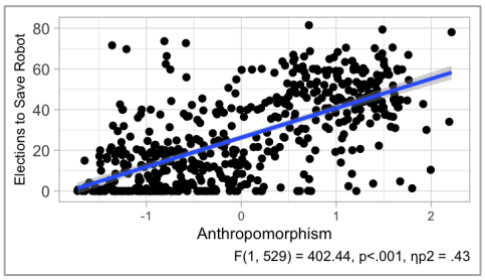

Figure 7: Anthropomorphism vs. Moral Regard for Robot.

Predicting Expectations of Capabilities A simple linear regression analysis revealed the factor Anthropomorphism predicts expectations of anthropomorphic capabilities, $\eta_p^2 = .74$ (Figure 6). This result and remarkably large effect size suggest the association between anthropomorphism and the tendency to overestimate social or emotional capabilities that the robot may not have (e.g. "Understand my emotions," "Give me advice") is important and merits further study.

Predicting Moral Judgements with Anthropomorphism. A simple linear regression analysis revealed the factor Anthropomorphism predicts elections to save the robot at greater risk of human safety in hypothetical moral dilemma scenarios, $\eta_p^2 = .43$ (Figure 7). The remarkably large size of this effect is striking and suggests that the association between anthropomorphism and moral care for the robot—even at greater risk to human life—can be affected even by relatively minimal manipulations of bottom-up and top-down cues to agency. This is an important topic for future research.

Discussion

Top-Down Effects on Anthropomorphism

The significant effect of description on anthropomorphism suggests that top-down influences on anthropomorphism matter. The hypothesis that the anthropomorphic description would cause more anthropomorphism relative to the neutral description is supported by these results. The anthropomorphic description also caused the use of more LMA (linguistic markers of anthropomorphism: considered a behavioral measure) and higher expectations of anthropomorphic capabilities. Taken together, these results suggest that personal pronouns, verbs that imply or refer to humanlike mental states, humanlike descriptors, and the robot's name—linguistic features that might occur in common discourse, product advertising, news reports, etc.—increase anthropomorphism in those exposed to them.

The hypothesis that the mechanistic description would decrease explicit anthropomorphism relative to the neutral description was not supported by these results. Those in the mechanistic and neutral description conditions did not anthropomorphize significantly differently; indeed, those in the mechanistic description group anthropomorphized slightly more than those in the neutral. Though unexpected, this result may be explained by a "sophistication bias": possibly, describing the robot's advanced technological capabilities promotes a conception of an advanced robot—and, critically, in the face of the "advanced robots" of present popular culture, the person may conceptualize it as also having qualities of a social, emotional entity. That said, the mechanistic description caused higher expectations of mechanistic capabilities and (though not significantly) the use of fewer LMA; both suggest a notable influence, and in particular the latter result suggests that those who hear the robot described mechanistically may be slightly less likely to describe the robot anthropomorphically in turn.

Bottom-Up and Top-Down Effects

The bottom-up effects shown in our results can be characterized by the influence of watching the way in which a human and robot interact. The Social Interaction video increased anthropomorphism and the use of LMA. The Mechanistic Interaction video decreased the use of LMA and

increased expectations of mechanistic capabilities relative to all other video conditions except No Video (which also featured high expectations). This pattern suggests that watching the way in which other humans interact with a robot—as if the robot were a tool, the robot primarily showing its tool functionalities; or as if the robot were a social companion capable of emotion, the robot appearing to look and behave accordingly—can influence anthropomorphism. The importance of this result is underscored by the fact that the bottom-up exposure of robots to most of the general population today occurs not through direct, co-present interaction with the robot, but through images and videos of the robot by itself or interacting with another human: this result suggests that those videos have a significant influence on the viewer's anthropomorphism of the robot.

Strikingly, the top-down effect of description had a larger influence on anthropomorphism (compared to bottom-up cues) and was the only manipulation to significantly influence the attribution of agency and experience capacities. These top-down cues, then, had the primary impact on causing participants to view the robot similarly to how they view adult humans (cf. Gray, et al., 2007). Though most research on anthropomorphism to date has focused on bottom-up cues, these results illustrate the equal and potentially greater importance of top-down cues.

Lastly, the broad lack of interactions of the effects of description and video suggests the top-down and bottom-up effects on anthropomorphism (and advanced tool perceptions) are relatively independent. While further research should continue to examine this hypothesis, these results suggest that top-down effects on anthropomorphism may occur largely independently from bottom-up effects.

Consequences of Conceptualizations

There is also evidence that anthropomorphism has important consequences. In line with the theorizing of Duffy (2003) and Scheutz (2012), our empirical results support the hypothesis that anthropomorphism promotes overestimations of the robot's social and emotional capabilities. That is, anthropomorphic cues—including top-down cues—may increase risks related to unrealistic expectations of a robot's actual functionalities. This finding has implications for human-robot cooperation, which requires an accurate understanding of capabilities, and contexts ranging from education and professional training to product advertisement.

Equally notably, and in line with Nijssen et al. (2019), our results suggest that anthropomorphism is associated with increased moral regard for the robot: the more one anthropomorphizes the robot, the more one is likely to elect to save the robot, even at greater risk to human safety. This result, if borne out in non-hypothetical situations, bears obvious ramifications important to consider when designing, marketing, or discussing robots that interact with humans. These associated consequences of anthropomorphism underscore the importance of seriously considering the top-down effects of the language used to describe robots.

Limitations

Critically, our results give no indication of the *persistence* or *generality* of the demonstrated effects: there remain important questions about how long a description will influence one's conception of the described robot, or whether the effect will influence one's conceptualizations of other robots. These questions merit further research.

Moreover, important questions remain about how the relative influences of top-down and bottom-up cues change (a) in an in-person context and (b) when the participant is interacting with the robot directly, rather than watching another human interact with the robot. There is some evidence that the effect of linguistic top-down cues lessens in an in-person context when the participant is interacting with a humanoid robot directly (Onnasch & Roesler, 2019; Wallkötter et al., 2020). With regard to these questions, it is a limitation of this study that the participants were exposed to the robot's bottom-up cues only in video format and not in an embodied co-present manner, and that the participant watched someone else interact with the robot rather than interacting with the robot directly. That said, the top-down and bottom-up stimuli of this study are realistic to online news reports, social media, and product advertisements—the real-world contexts in which many (and, at present, most) people are actually exposed to robots, in which the viewer does not interact with the robot directly, but rather sees the robot by itself or interacting with another human, along with some linguistic description. Thus, though the present study is highly limited in answering questions about co-present direct human-robot interaction, this study's results are still of high importance due to its reasonable degree of realism to the context in which most will be exposed to robots today—and the implications are critical for the opinions developed about robots and the roles we give them in the days to come.

Conclusion

The results of this study suggest that top-down effects on anthropomorphism matter. Using different language to describe a robot can influence the degree to which those exposed to those words anthropomorphize it—which, in turn, has critical downstream consequences that can influence the place we give robots in our social and ethical frameworks. Further, using this language seems to cause others to be more likely to use similar language, which will in turn influence still others' conceptions and behavior.

As robots and other machines become more integrated into our daily lives, a more thorough understanding of the causes and consequences of anthropomorphism becomes more critical. Top-down effects on anthropomorphism seem to matter for our thought, talk, and potential treatment of robots. Especially as we set norms for discussing robots in common discourse, education, marketing, media, and policy, top-down effects merit serious consideration and further study.

Supplemental Materials, Data, and Code

https://doi.org/10.7910/DVN/B1O1ZQ

References

Anki. (2019). Vector. Anki. <https://anki.com/en-us/vector.html.>

Bryson, J. J. (2010). Robots should be slaves. *Close Engagements with Artificial Companions: Key social, psychological, ethical and design issues*, 63-74.

Darling, K. (2017). "'Who's Jonny?' Anthropomorphic Framing in Human-Robot Interaction, Integration, and Policy" in P. Lin, R. Jenkins, & K. Abney (Eds.), *Robot Ethics 2.0* (173- 188). New York: Oxford University Press.

Darling, K., Nandy, P., & Breazeal, C. (2015). Empathic concern and the effect of stories in human-robot interaction. In *24th IEEE International Symposium on Robot and Human Interactive Communication*, 770-775). IEEE.

Duffy, B. R. (2003). Anthropomorphism and the social robot. *Robotics and Autonomous Systems*, *42*, 177-190.

Eyssel F, Kuchenbrandt D. (2012). Social categorization of social robots: anthropomorphism as a function of robot group membership. Br. J. Soc. Psychol. 51:724–31.

Forlizzi, J. (2007). How robotic products become social products: an ethnographic study of cleaning in the home. In *2007 2nd ACM/IEEE International Conference on Human- Robot Interaction (HRI)* (pp. 129-136). IEEE.

Fussell, S. R., Kiesler, S., Setlock, L. D., & Yew, V. (2008). How people anthropomorphize robots. In *2008 3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 145-152). IEEE.

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science*, *315*(5812), 619-619.

Gray, K., & Wegner, D. M. (2012). Feeling robots and human zombies: Mind perception and the uncanny valley. *Cognition*, *125*(1), 125-130.

Haring, K. S., Watanabe, K., & Mougenot, C. (2013, March). The influence of robot appearance on assessment. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 131-132). IEEE.

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *The American Journal of Psychology*, *57*(2), 243-259.

Johnson, S. C. (2003). Detecting agents. *Philosophical Transactions of the Royal Society of London*, *358*, 549-559.

Matthews, G., Lin, J., Panganiban, A. R., & Long, M. D. (2019). Individual Differences in Trust in Autonomous Robots: Implications for Transparency. *IEEE Transactions on Human-Machine Systems*, 1-11.

Mitchell, R. L. C., & Xu, Y. (2015). What is the value of embedding artificial emotional prosody in human-computer interactions? Implications for theory and design in psychological science. *Frontiers in Psychology*, *6*, 1-7.

Nijssen, S. R., Müller, B. C., Baaren, R. B. V., & Paulus, M. (2019). Saving the robot or the human? Robots who feel deserve moral care. *Social Cognition*, *37*(1), 41-S2.

Novikova, J., & Watts, L. (2015). Towards artificial emotions to assist social coordination in HRI. *International Journal of Social Robotics*, 7(1), 77-88.

Onnasch, L., & Roesler, E. (2019, November). Anthropomorphizing robots: The effect of framing in human-robot collaboration. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 63, No. 1, pp. 1311-1315). Sage CA: Los Angeles, CA: SAGE Publications.

O'Neal, A. L. (2018). *Is Google Duplex Too Human? Exploring User Perceptions of Opaque Conversational Agents* (Unpublished master's thesis). University of Texas at Austin, Austin.

Richards, N. M. & Smart, W. D. (2016). "How Should the Law Think about Robots?" in R. Calo, M. Froomkin, & I. Kerr (Eds.), *Robot Law* (3-24). Cheltenham: Edward Elgar.

Scheutz, M. (2012). "The Inherent Dangers of Unidirectional Emotional Bonds between Humans and Social Robots" in P. Lin, K. Abney, & G. A. Bekey (Eds.), *Robot Ethics* (205-222). New York: Oxford University Press.

Schulevitz, J. (2018). Alexa, Should We Trust You? *The Atlantic*.<https://www.theatlantic.com/magazine/archive/2018/11/alexa-how-will-you-change- us/570844/>.

Shellenbarger, S. (2019). Why We Should Teach Kids to Call the Robot 'It.' *The Wall Street Journal*. < https://www.wsj.com/articles/why-kids-should-call-the-robot-it-11566811801>.

Stone, Z. (2017). Everything You Need To Know About Sophia, The World's First Robot Citizen. *Forbes*. < https://www.forbes.com/sites/zarastone/2017/11/07/everything-you-need-to-know-about-sophia-the-worlds-first-robot-citizen/?sh=6723a90f46fa>.

Thellman, S., Silvervarg, A., & Ziemke, T. (2017). Folk-psychological interpretation of human vs. humanoid robot behavior: Exploring the intentional stance toward robots. *Frontiers in psychology*, *8*.

Wada, K., Kouzuki, Y., & Inoue, K. (2012, June). Field test of manual for robot therapy using seal robot. In 2012 4th IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob) (pp. 234-239). IEEE.

Wallkötter, S., Stower, R., Kappas, A., & Castellano, G. (2020, March). A Robot by Any Other Frame: Framing and Behaviour Influence Mind Perception in Virtual but not Real-World Environments. In Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (pp. 609-618).

Waytz, A., Heafner, J., & Epley, N. (2014). The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology*, *52*, 113-117.