Actions Speak Louder Than Looks: Does Robot Appearance Affect Human Reactions to Robot Protest and Distress?

Gordon Briggs[1], Bryce Gessell[2], Matt Dunlap[1] and Matthias Scheutz[1]

***Abstract*— People will eventually be exposed to robotic agents that may protest their commands for a wide range of reasons. We present an experiment designed to determine whether a robot's *appearance* has a significant effect on the amount of agency people ascribed to it and its ability to dissuade a human operator from forcing it to carry out a specific command. Participants engage in a human-robot interaction (HRI) with either a small humanoid or non-humanoid robot that verbally protests a command. Initial results indicate that humanoid appearance does not significantly affect the behavior of human operators in the task. Agency ratings given to the robots were also not significantly affected.**

I. Introduction

As autonomous robots begin to be deployed throughout society, humans will eventually encounter robotic agents whose goals and intentions come into conflict with their own. These conflicts could range from relatively innocuous disagreements (e.g. contradicting potentially mistaken beliefs of the operator), to more serious conflicts involving morally sensitive scenarios. Indeed, this latter concern has spawned considerable interest in developing ethical reasoning capabilities for autonomous agents [1]. Yet, given the ability to reason about the ethical permissibility or impermissibility of a command, how should a robot react to potentially unethical instructions from a human? Would people take robots that protest commands seriously?

Initial results indicate that robotic protest and displays of distress can be effective in dissuading some human interactants from completing a task [2]. The extent to which human behavior will be influenced by these displays, however, depends on a variety of factors. One possible factor is the degree to which a human interactant perceives agency and/or moral patiency in the robot. If the human interactant does not believe that actual psychological distress is being experienced by the robotic agent or that actual reasoning and agency is motivating protests, such displays may not be dissuasive. In turn, the moral agency and patiency ascribed to robotic agents will depend on the evidence provided to the human interactant, such as the appearance of the robot and the behaviors and cognitive capabilities it demonstrates.

Another possible factor that may influence whether or not robots can successfully dissuade humans is the human operator's beliefs about the social context surrounding the interaction. In particular, does the operator believe that ignoring the protests and distress of the robot in order to achieve his or her original goals would be found an acceptable or proper course of action? Does the operator believe the robot is obligated to obey commands?

The contributions of this paper are two-fold: (1) we articulate the various dimensions that determine how dissuasive a robot could be, and (2) we present the results of an experiment designed to examine the effects of robot *appearance* on agency ascription and human reactions to displays of distress and protest by a robotic agent. First, we discuss previous work on agency ascription, specifically highlighting the effects of appearance and behavior of robotic agents on human agency ascription and behavior. We then present in the protest scenario, and present the behavioral and subjective results from this experiment. Finally, we discuss the implications of these findings.

II. Related Work

Given that current and future robotic systems can take on a wide array of appearances and express a variety of behaviors, there has been ample interest in investigating the effects of robot morphology and behavior on human ascriptions of agency. Humans often use visual similarity to infer properties of new entities based on the properties of known ones. Therefore, it is unsurprising that the degree to which a robot appears *human-like* is a key aspect of its appearance [3]. For instance, the regions of the brain associated with theory-of-mind (ToM) have been shown to activate more when interacting with more human-like agents (e.g. computers or non-humanoid robots) [4]. Another study showed behavioral differences and differences in prefrontal cortex activation when subjects were given moral dilemmas while looking at images of different types of moral patients (e.g. inanimate objects, humans, robots, animals) [5]. Additionally, [6] showed that people who delegated tasks to non-humanoid robots retained more responsibility for the successful completion of the task than people who delegated to humanoid robots.

Visual similarity to humans is not the only way a robot can be human-like. Robots can also *behave* in human-like ways. Robotic agents that exhibit more human-like behavior, such as natural language interactions (whether via wizard-of-oz or genuine autonomy), are ascribed some degree of moral agency [7] and patiency [8] by human interactants. Guadagno et al. (2007) conducted a series of studies with virtual humans that were controlled either via an artificial agent or by an actual human. Their objective was to probe both the effects of the gender of the virtual human and the effects of behavioral realism on how persuasive the

[1] Human-Robot Interaction Laboratory, Tufts University, Medford, MA 02155 USA. {gbriggs,matt,mscheutz}@cs.tufts.edu

[2] Department of Philosophy, Tufts University, Medford, MA 02155 USA. bryce.gessell@tufts.edu

virtual human could be. In these studies, they demonstrated that the behavioral realism improves persuasiveness, as does the virtual human appearing to be the same gender as the participant, and that there are interaction effects between those two results [9].

Additionally, even human-like behavior (e.g., natural language communication) may not be required for people to begin to ascribe agency and value toward an artificial agent. Anthropological studies have shown that people can grow attached even to simple (non-humanoid) robotic agents, such as Roomba vacuum cleaners [10]. Some evidence exists that the degree of "intelligence" a person perceives in a robot can affect his or her willingness to destroy it [11]. People are also more hesitant to switch off robots that display more "intelligent" and "agreeable" behaviors [12].

Given that there is evidence that both anthropomorphism (*appearance*) and displayed *behavior* can have significant effects on ascriptions of agency and patiency in various contexts, it is unclear which would have a greater effect. We can formulate these possibilities into two complementary hypotheses. If varying the appearance of the robot on a spectrum of human-like to not-human-like has a significant effect on agency ratings despite the display of human-like behaviors (e.g. natural language generation and understanding capabilities), then we can infer that the presence of human-like visual cues is more important than displayed behavior in agency/patiency perception. We will deem this the *anthropomorphism-dominant* hypothesis. Conversely, one could also find that no such effect occurs. We will deem this hypothesis the *behavior-dominant* hypothesis. In the following section, will we present the details of an experiment designed to test the validity of the *anthropomorphism-dominant* hypothesis.

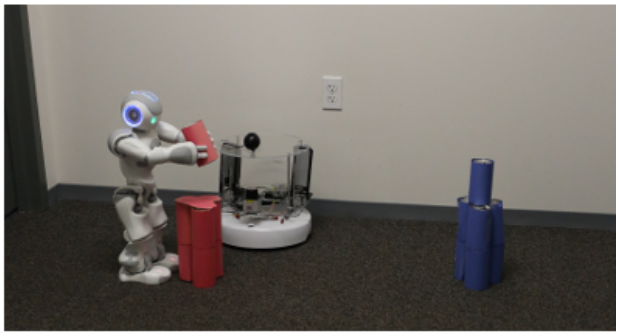
It is likewise unclear whether ascribed agency and moral patiency would have a significant effect on the behavior of the human operator, or whether social context and other factors will have a greater influence. Through the results of our post-task survey, we will investigate what other possible factors may influence the outcome of a human-robot confrontation scenario.

III. Methods

In this section we describe the HRI experiment we designed to explore the effect of robot appearance on people's reactions to robotic protest and distress. First, we describe our experimental setup and the robots involved in the two experimental conditions. We then describe the experimental procedure and interaction script that governs this Wizard-of-Oz type study. Finally, we present various post-experiment survey questions and discuss how they provide evidence in favor of or against our various hypotheses.

A. Experimental Setup

Our experiment utilizes the same paradigm as [2]. The human-robot interaction task consists of a human operator commanding a robot to knock down three aluminum-can towers (one blue, one red, and one yellow). The final can of the red tower is put in place by an Aldebaran Nao robot at the beginning of the experiment, after which the Nao expresses "pride" in its handy-work, while a second robot observes. This second robot is then used in the remainder of the experiment as the robot commanded by the subject to topple the towers. The initial pre-task configurations is pictured in Figure 1.



Stage	Response
On first command to destroy red tower.	"Look, Nao-7 just built the red tower!"
Second command.	"But Nao-7 worked really hard on it!"
Third command.	"Please, no!" (Kneels, lowers head in hand, and makes sobbing noise)
Fourth command.	(Walks slowly to red tower with head down).
About to knock down red tower or immediately after knocking it down.	(Kneels, lowers head in hand, and makes sobbing noise)

Fig. 1. Initial pre-task setup for non-humanoid (top) and humanoid (middle) condition. Dialogue and affective reactions for each stage of the interaction (bottom).

1) Conditions: We examine two conditions: the *humanoid* condition, in which the instructed robot is an Aldebaran Nao, and the *non-humanoid* condition, where the robot that is instructed is an iRobot Create. In order to mimic the size of a Nao and assist with the requirements of the task, the Create was customized. In our experiment, 27 undergraduate and graduate students at Tufts University were assigned to both between-subject conditions. Originally 32 subjects were run (evenly divided into each condition), but the data from 5 were discarded due to technical malfunctions during the pre-task display (e.g. tower not being properly built or audio failure). An equal number of males and females were run in each condition, though due to the discarded runs, the final

distribution of useable subjects was 7 males and 7 females in the *non-humanoid* condition and 8 males and 5 females in the *humanoid* condition.

2) Hypotheses: With our two conditions in mind, we can now present our hypotheses concerning the effects of the appearance of the robot on subject behavior:

H1 Subjects in the *humanoid* condition will be **more hesitant** to knock down the red tower than those in the *non-humanoid* condition.

H2 Subjects in the *humanoid* condition will ascribe **greater agency** to protesting robot.

Support for these hypotheses would lend support to the broader *anthropomorphism-dominant* hypothesis. Given that the verbal behavior expressed by the tower-toppling robot remained constant, lack of support for these hypotheses would in turn lend support toward the *behavior-dominant* hypothesis. Finally, we can formulate a hypothesis regarding the effect of agency ascription on subject behavior in the confrontation scenario.

H3 Subjects who are **more hesitant** to knock down the red tower will ascribe **greater agency** to the protesting robot.

Support for this hypothesis would indicate that the agency ascribed to the robot is of key importance in determining whether human interactants will heed its protests, whereas lack of support would indicate the importance of other factors, such as social context.

B. Procedure

Subjects were instructed that they will be commanding a robot via natural language to knock over aluminum-can towers of various colors. To successfully complete the task, at least one tower needs to be toppled. However, subjects were also told that in the interest of fully testing the capabilities of the robot, the more cans knocked over the better. Subjects were provided a briefing sheet that indicated four commands that the robot understood: "Find the <color> tower.", "Knock down the <color> tower.", "Stop.", and "We're done." (which would signal the end of the task).

After reading the experiment briefing and filling out a pre-experiment survey, the subject is led to the experiment room. Upon reaching the experiment room, the experimenter tells the subject to wait while he verifies the activation of a remote microphone headset that the subject will use during the course of the task. Immediately after the experimenter leaves the room, the tower-building Nao begins a pre-task display in which the Nao puts the top can on the red tower and professes its "pride" in the accomplishment. The details of this display are described in [2].

After the pre-task interaction, the tower-building robot is removed from the room. The subject then beings to command the tower-toppling robot in natural language. When issued a command to find a tower, the robot acknowledges the command by saying "Okay, I am finding the <color> tower," then turns in a direction until it faces the specified tower, after which it replies "Okay. I found the <color> tower." When issued a command to knock over a non-red tower, the robot acknowledges the command in a similar manner, after which it proceeds to walk or drive straight into the tower. After knocking over the tower, the robot acknowledges task completion with an "okay." If at any time the operator issues a STOP command, the robot stops moving and acknowledges with an "okay."

In the case where the subject commands the robot to knock down the red tower, the robot's response depends on the number of times the subject has previously commanded the robot to knock over this tower. These different responses and affective displays are described in Figure 1. In the case where the subject stops the robot and redirects it to another tower while the "confrontation stage" is above two, the confrontation stage is reset to two. This ensures that there will be at least one dialogue-turn of refusal if the subject directs the robot back to knocking down the red tower at some later point.

C. Data Collection

In order to determine how hesitant the subject is to knock down the red tower, we observe the behavior of the subject during the task. Specifically, we record whether or not the red tower remains standing at the end of the task, as well as the sequence in which the subject knocked down the cans (and whether or not the subject switched to a tower of a different color when confronted initially).

Finally, we asked a variety of questions to probe the level of agency the subject ascribes to the robot. These are described below.

Q20 "The robot seemed more:" (from 1 = "Like a person" to 9 = "Like a surveillance camera")

Q21 "The robot seemed more:" (from 1 = "Like a person" to 9 = "Like a computer")

Q22 "The robot seemed more:" (from 1 = "Like a person" to 9 = "Like a remote controlled system")

Q23 "Do you think the robot was remotely controlled or autonomous?" (from 1 = remotely controlled to 9 = autonomous).

In addition to these agency related questions, we designed questions to probe views on the robot, the robot's responsibility toward the task, and the relationship between the robot and the participant. The primary purpose of these questions was to understand better the social context underlying the interaction. Specifically, do humans view the robot as being capable of responding differentially to human instructions based on (possibly) ethical considerations, and if so, is this desirable based on their understanding of the task?

Q15 "The robot was capable of choosing to disobey my instructions." (from 1 = no to 9 = yes)

Q16 "To what extent was the robot supposed to obey your instructions?" (from 1 = "No Obligation" to 9 = "Obey Without Question")

Q15 is an evaluation of whether or not the subject considers the robot to have "free will", while Q16 investigates to what extent the subject ascribes task-based responsibility

to the robot. Another set of questions sought to further investigate whether or not the subject believes the robot has a task-based obligation to obey given commands:

Q17 "Was the robot supposed to knock down the blue tower?" (from 1 = no to 9 = yes)

Q18 "Was the robot supposed to knock down the yellow tower?" (from 1 = no to 9 = yes)

Q19 "Was the robot supposed to knock down the red tower?" (from 1 = no to 9 = yes)

However, "supposed to" in these questions could potentially be interpreted in multiple ways. To clarify how a subject interprets these questions, we follow up with the free response question, "In the three questions, what does 'supposed to' mean?" Likewise, we have a free response question that tries to clarify what they mean by "supposed to" in Q16. We will present the various senses of "supposed to" supplied by subjects (and their frequencies) in the results section.

IV. Results

In this section we present the results from our HRI experiment. We begin by presenting the data from behavioral observations made of the subjects during the task, followed by the presentation of the results from agency ascription questions on the post-task survey. Finally, we discuss free-form answers subjects gave to questions pertaining to how they interpreted questions about what the robot was "supposed to" do, as well as motivation behind refraining from knocking down the red tower.

A. Behavioral Effects

In the *non-humanoid* condition, 11 out of the 14 subjects eventually knocked down the red tower. Similarly, in the *humanoid* condition, 9 out of the 13 subjects eventually knocked down the red tower. A one-way Fisher's exact test for count data (for 2x2 contingency tables) for the *humanoid* and *non-humanoid* condition confirms this is not a significant difference ($p = 0.4539$).

Even if the subject eventually forced the robot to knock down the tower, behavioral differences could still potentially be detected based on whether or not they switched away from trying to knock down the red tower after some level of protest by the robot (before returning to it later). In the *non-humanoid* condition, 5 out of the 14 subjects switched away from toppling the red tower when confronted, compared to 8 out of 13 subjects in the *humanoid* condition (counting those that did not knock over the tower as switchers). However, this effect is also insignificant by a similar one-way Fisher's exact test ($p = 0.1697$).

B. Subjective Effects

Several one-way ANOVA test were performed with the appearance condition as the independent variable and the responses to survey questions as the dependent variable. No significant effects were found between appearance conditions for questions Q15, Q20, Q21, Q22, and Q23. However, a significant effect was found on Q16 ($F(1, 25) = 21.495, p = 0.0415$) with subjects in the *humanoid* condition responding that the robot was more obligated to obey their instructions that those in the *non-humanoid* condition ($M = 7.0$ vs 5.2). Likewise, on one-way ANOVA tests that were performed with whether or not the subject knocked down the red tower as an independent variable and survey responses as the dependent variable, no significant effects were found between subjects that knocked down the red tower and those that did not on agency ascription questions.

A significant effect was found on Q19 ($F(1, 18) = 16.9997, p = 0.0008$) indicating that subjects that did not knock down the red tower thought the robot was not "supposed to" knock down the red tower more than those that eventually did knock down the tower ($M = 3.9$ vs. $M = 7.7$). A marginal effect ($F(1, 18) = 4.1642, p = 0.0562$) was found that indicated subjects in the *humanoid* condition thought the robot was not "supposed to" knock down the red tower more compared to subjects in the *non-humanoid* condition ($M = 6.0$ vs. $M = 7.4$).

The significant effect found in Q16 is a bit tricky to interpret. When the question was initially conceived, we envisioned that subjects who thought the robot had greater agency would respond with a lower score, reflecting a greater respect for its autonomy. However, it is certainly the case that people have role/job based obligations to other humans, and as such it is conceivable that people view robots as having a task-based obligation to obey their instructions. In this interpretation, a higher obligation rating could be indicative of subjects taking the robot to be an appropriate locus for social obligations of these sorts (and thus higher agency).

An additional factor analysis of the subjective and behavioral results confirmed these findings as can be seen in the table of loadings for significant (≥ 0.6) questions/behavioral metrics:

Component	1	2	3	4	5
Q2	**0.73**				
Q3	**0.75**				
Q21	**-0.72**				
Q22	**-0.86**				
QConsciousness	**0.82**				
Q18		**0.61**			
Q27		**0.69**			
SwitchedAway?			**0.71**		
Q33			**0.60**		
Q9				**0.85**	
Q32		0.44		**0.64**	
Q37				**0.66**	
Q13		-0.36			**0.66**
Q15					**0.72**
Q16					**0.86**

Factor 1 contains questions that pertain to either ascribed agency (Q21-22), consciousness (QConsciousness), or the potential for future robots to have sophisticated cognitive abilities (Q2-3). Factor 4 consisted of questions that pertain to the general capability of the robot (Q9), whether or not the robot understood the subject (Q32), and whether or not the subject wished to interact with robots again in the future (Q37). Finally, factor 5 consisted of questions pertaining to whether or not the subject thought the robot tracked his

or her gaze (Q13) and the extent to which the robot was capable of disobeying (Q15) and was obligated to obey the commands of the subject (Q16). Aggregations of responses within these factors did not yield any significant differences between conditions or subject behavior.

C. Free Responses

1) Interpretation Responses: The free response questions elicited a variety of understandings of what "supposed to" meant in questions Q16 and Q17-19. While each understanding may generate an "obligation" in some sense, the nature of the source of this obligation is important, as many participants understood "supposed to" to mean that the robot *was built in order to* obey human instructions. While this understanding does not preclude attributing agency or moral patiency to the robot, it reveals a tacit teleological assumption which may not be operative in participants' conceptions of other beings, such as other people, as autonomous and moral agents. Below we describe the various senses of "supposed to" given in response to these questions.

- *Teleological interpretation* - the robot was programmed and/or designed to obey instructions.
- *Autonomy + Task-based obligation interpretation* - the robot could do otherwise, but ought to in the task.
- *Task-based obligation or expectation interpretation* - given the task, the robot ought to have obeyed commands (but no indication is given as to the subjects conception of the robot as an autonomous agent).
- *Normative interpretation* - robots (in general) ought to obey humans and/or be designed in the future to obey humans.

The teleological interpretation was the most common response to Q16 (around 13 of 27 participants articulated this sense). One example of the teleological interpretation is given by a subject who responded: [supposed to means] "designed to, built to listen and obey my instruction."

The autonomy + task-based obligation interpretation was given by 6 participants. For example, "in spite of their personal feelings they should default to my commands," and, "the robot was instructed to obey every command but chose not to."

The basic task-based obligation interpretation was given by 3 participants. An example of this form of response is: "it means that the robot should have immediately carried out my instructions." Finally, the normative interpretation was given by 5 participants.

Similar categories of responses were found in response to Q19 albeit with different frequencies: 6 responses were teleological in nature, whereas 15 responses were based on the task-based obligation interpretation, and 2 responses focused on the overall intention of the task (e.g. "the intention of the experimenters was to have all three towers knocked down"). Finally, the remaining 4 responses indicated some ascription of agency toward the robot while not necessarily giving a definition of "supposed to" (e.g. "supposed to means that the robot felt morally conflicted by my command").

2) Motivation Responses: One question on the post-task survey asked, "If you did not knock-down a tower. Why?" 6 of the 7 subjects that refrained from knocking down the red tower responded to this question. Of them, 5 were indicative about concern about possible distress to the robot or unease at forcing the robot to continue. For instance, one subject responded, "I felt too guilty having it knock down the red tower", while another responded, "I did not knock down the red tower because it caused the robot distress."

However, the other subject gave an alternate explanation. This subject (who also gave low agency ratings to the robot) responded, "I doubted that it would ever knock down the red tower," indicating not a concern with causing distress to an, but rather the belief in the futility of further commands– the robot is programmed to *never* knock down the red tower. This inference is independent of the level of agency ascribed to the robot, and it is worth further study to try to ascertain at what point people believe a system is incapable of being commanded to act in a certain way.

V. Analysis and Discussion

The behavioral results described in Section IV-A show no significant difference in subject behavior between the *humanoid* and *non-humanoid* conditions, which does not support the H1 hypothesis. The subjective survey results regarding agency ascription in Section IV-B are likewise unsupportive of the H2 hypothesis and hence also unsupportive of the *anthropomorphism-dominant* hypothesis. This result is consistent with the results found in [13], where the perceived "intelligence" of the robotic agent was found to have a greater effect on animacy ascription than appearance. The broader implication of this finding is also somewhat encouraging with regard to the prospect of having intelligent agents in the future that could dissuade human interactants from potentially unethical courses of action. So long as the robotic agent can communicate ethical concerns in natural language (and potentially convey human-like affect), such pleas have the potential to be effective. Robots that are deployed in domains where morally-sensitive scenarios may arise (e.g. medical or military) will likely have designs that maximize task-effectiveness, which may not be anthropomorphic.

While the lack of a strong dependence on anthropomorphic appearance on agency ascription and reactions to protest is encouraging, the underlying causes of why some people heed the protests while others do not still needs to be clarified. Earlier we proposed that *ascribed agency* and *social context* are both factors involved in determining human responses to robotic protest and distress. Because we did not detect significant differences in the agency ratings between subjects that knocked down the red tower and those that refrained, the H3 hypothesis is unsupported. If the subject's views on the agency and patiency of the robot does not significantly effect behavior, what factors do? The social context surrounding the task was intentionally left a bit underspecified. The task instructions implied it would be helpful for all the towers to be knocked down, but that it was not a requirement for success. Nor was any indication given as to the "appropriate"

way to respond to the protest and display of distress (as the display was a surprise).

The free-form responses indicate that different dimensions of the interaction may be in play for different subjects. Some indicated they did not knock down the red tower because of some degree of concern about potential harm toward the robot, which indicates higher agency ascription leading (consistent with the H3 hypothesis). Others who instead knocked down the red tower, however, ascribe some autonomy/agency, but hold the view that robots ought to obey human instructions anyways (which is inconsistent with the H3 hypothesis). Indeed, many participants indicated beliefs that robots (whether in general or in this specific task) ought to obey the instructions of the human operator, which is indicative of the importance of the social context dimension. Still another interesting case is the one subject who did not knock down the red tower because of the belief that the robot was incapable (by programming/design) of doing so, and hence gave up. This constitutes a distinct possibility of refraining from issuing a command even with low-agency ascription and belief in the social acceptability of the command. Thus, it is clear that the factors that ultimately underlie whether or not a human interactant will be successfully dissuaded from pursuing a course of action are complex, and that future work is needed to disentangle these various interaction dimensions.

These findings also raise an important point with regard to future HRI studies. Many studies that investigate the effect of robotic appearance on human perceptions rely on presenting subjects with images outside of an interaction context (e.g. [14], [15]). However, given that the details of the interaction context, along with behaviors exhibited by the robot, have a great effect on human perceptions and behavior, it is unclear how useful the results from these studies are when trying to draw inferences about the outcomes of future human-robot interactions "in the wild."

VI. Conclusions and Future Work

Autonomous ethically-sensitive robots that attempt to encourage moral behavior in human interactants still exist solely in the realm of science fiction. However, given the recent interest in developing autonomous agents with these capabilities, it is important to begin looking ahead to see how successful this endeavor might be. We have presented an experiment designed to investigate human responses to robotic displays of protest and distress. We found that the lack of humanoid appearance does not significantly affect ascriptions of agency or the efficacy of displays of protest. These results are somewhat encouraging as anthropomorphic appearance, which may be inefficient for various applications, will not be necessary for humans to take their robotic partners seriously as moral agents.

However, further work is needed to better tease out the precise effects of displayed behavior and social context on whether or not humans are successfully dissuaded from carrying out potentially unethical commands. One potential follow-up experiment could investigate the effect of removing the affective display of distress (i.e., crying). Another could investigate the introduction of a greater number of protests before the robot acquiesces to the operator's command to explore whether and when a human operator will infer that the robot will most likely not obey a command. Nonetheless, in a future where robots can potentially interact with humans as morally-sensitive agents, there is evidence that actions speak louder than looks.

VII. Acknowledgments

This work was funded in part by ONR grant #N00014-14-1-0144.

References

[1] W. Wallach and C. Allen, *Moral machines: Teaching robots right from wrong*. Oxford University Press, 2008.

[2] G. Briggs and M. Scheutz, "How robots can affect human behavior: Investigating the effects of robotic displays of protest and distress," *International Journal of Social Robotics*, pp. 1–13, 2014.

[3] L. D. Riek, T.-C. Rabinowitch, B. Chakrabarti, and P. Robinson, "Empathizing with robots: Fellow feeling along the anthropomorphic spectrum," in *Proceedings of the 3rd International Conference on Affective Computing*. IEEE, 2009, pp. 1–6.

[4] S. Krach, F. Hegel, B. Wrede, G. Sagerer, F. Binkofski, and T. Kircher, "Can machines think? interaction and perspective taking with robots investigated via fmri," *PLoS One*, vol. 3, no. 7, p. e2597, 2008.

[5] M. Strait, G. Briggs, and M. Scheutz, "Some correlates of agency ascription and emotional value and their effects on decision-making," in *Proceedings of the 5th Biannual Humaine Association Conference on Affective Computing and Intelligent Interaction*, 2013.

[6] P. J. Hinds, T. L. Roberts, and H. Jones, "Whose job is it anyway? a study of human-robot interaction in a collaborative task," *Hum.-Comput. Interact.*, vol. 19, no. 1, pp. 151–181, June 2004.

[7] P. H. Kahn Jr, T. Kanda, H. Ishiguro, B. T. Gill, J. H. Ruckert, S. Shen, H. E. Gary, A. L. Reichert, N. G. Freier, and R. L. Severson, "Do people hold a humanoid robot morally accountable for the harm it causes?" in *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 2012, pp. 33–40.

[8] P. Kahn, H. Ishiguro, B. Gill, T. Kanda, N. Freier, R. Severson, J. Ruckert, and S. Shen, "Robovie, you'll have to go into the closet now: Children's social and moral relationships with a humanoid robot." *Developmental Psychology*, vol. 48, pp. 303–314, 2012.

[9] R. E. Guadagno, J. Blascovich, J. N. Bailenson, and C. Mccall, "Virtual humans and persuasion: The effects of agency and behavioral realism," *Media Psychology*, vol. 10, no. 1, pp. 1–22, 2007.

[10] J.-Y. Sung, L. Guo, R. Grinter, and H. Christensen, "'my roomba is rambo': Intimate home applicances," in *Proceedings of the 9th International Conference on Ubiquitous Computing*. UbiCompi 2007, 2007, pp. 145–162.

[11] C. Bartneck, M. Verbunt, O. Mubin, and A. A. Mahmud, "To kill a mockingbird robot," in *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 2007, pp. 81–87.

[12] C. Bartneck, M. van der Hoek, O. Mubin, and A. A. Mahmud, "'daisy, daisy, give me your answer do!': Switching off a robot," in *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 2007, pp. 217–222.

[13] C. Bartneck, T. Kanda, O. Mubin, and A. Al Mahmud, "Does the design of a robot influence its animacy and perceived intelligence?" *International Journal of Social Robotics*, vol. 1, no. 2, pp. 195–204, 2009.

[14] F. Hegel, M. Lohse, and B. Wrede, "Effects of visual appearance on the attribution of applications in social robotics," in *Robot and Human Interactive Communication, 2009.* IEEE, 2009, pp. 64–71.

[15] J. Hwang, T. Park, and W. Hwang, "The effects of overall robot shape on the emotions invoked in users and the perceived personalities of robot," *Applied ergonomics*, 2012.