

Overselling: Is Appearance or Behavior More Problematic?

Gordon Briggs
Human-Robot Interaction Laboratory
Tufts University
Medford, MA 02155, USA
gordon.briggs@tufts.edu

ABSTRACT

As robots become increasingly ubiquitous in society, ethical questions will arise based on discrepancies between the perceived and actual capabilities of these agents: some of which may connote some degree of moral patency or agency. In this abstract, we present the position that in some contexts the observed *behavior* of a robotic agent may be a more powerful determinant of the degree of agency and patency people ascribe to robotic agents, rather than mere *appearance*. If so, concerns regarding the mismatch between inferred and actual capabilities of robots (and their consequent moral status) should also be concerned with not overselling the behavior of robotic agents in addition to the morphology of robotic agents.

1. INTRODUCTION

As technologies that enable the deployment of increasingly autonomous robots into society are developed, the robotics and human-robot interaction (HRI) community will be faced with a variety of questions surrounding how people who interact with these artificial agents view these agents as moral patients (i.e. appropriate targets of moral considerations and/or social and emotional investment). These questions will, in turn, lead to a variety of ethical concerns that derive from the potential mismatch between the level of patency ascribed toward robotic agents by naive users and the level of patency warranted by the various cognitive architectural mechanisms possessed by these agents (which, for the foreseeable future, will be rather low). For instance, some researchers have expressed concerns about the development of one-way social relationships, in which people become emotionally invested in robots that have no means with which to *actually* reciprocate this attachment or benefit from it [5].

In order to attempt to mitigate this potential mismatch, it is necessary to begin to understand the psychological processes that are at work when human users make inferences about the capacities of the robotic agents with which they are interacting, and how these inferences affect the degree of

patency and agency attributed to these robots. At a high-level, the process of inferring the capacities of an agent is the process of weighing the evidence for particular capacities based on observable cues.

For instance, possessing an arm may provide evidence that one has the ability to grab and manipulate objects. In comparison, possessing the ability to hold (for a limited time) an intelligent conversation with references to past interactions may connote a variety of sophisticated cognitive and social capabilities, which may or may not be present. Given this formulation of the problem, there are a couple key questions that must be addressed by future research:

- What set of inferred capabilities are associated with ascriptions of moral patency and agency?
- What happens when multiple observations provide contradictory conclusions?

In this abstract, we present the position that in many contexts the observed *behavior* of a robotic agent may be a more powerful determinant of the degree of agency and patency people ascribe to robotic agents, rather than mere *appearance*. As such, concerns regarding the mismatch between inferred capabilities and moral status of robots and reality should also be concerned with not overselling the behavior of robotic agents as well as not overselling the morphology of robotic agents.

2. MORPHOLOGY VS. BEHAVIOR

There is, of course, no question that people can and will infer some degree of human-like capabilities based on the evidence presented by the appearance of an agent. Inferring characteristics of entities based on visual similarity to other entities is a oft-used cognitive capacity by people, and there is empirical evidence that indeed people infer human-like characteristics in human-looking robots [4]. However, the limitation of many of these studies is that they either present static images of robots (lacking any behavioral evidence) or they present human-like robots in conjunction with human-like behavior. Could the inferences made by human-like behavior override the inferences made by the absence of human-like morphological cues? Or vice versa?

2.1 Experimental Evidence

We conducted a set of experiments designed to probe how human interactants reacted to robots that displayed (simulated) protest and distress behaviors toward a particular command during a collaborative task [2]. Using the

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author. Copyright is held by the owner/author(s).
The Emerging Policy and Ethics of Human Robot Interaction, Mar 2, 2015, Portland, OR, USA.

same paradigm, we investigated the question of whether or not humanoid or non-humanoid appearance would significantly affect reactions and/or agency ratings given toward the robot [1]. We found that non-humanoid appearance did not significantly affect either of these metrics. What was constant in both the humanoid and non-humanoid conditions, however, was rather human-like natural language behavior. The robot would (via a Wizard-of-Oz setup) be responsive to verbal instruction, and in the case of a protested command, would refuse based on justifications rooted in appeals to rather human-like mental states (e.g. not wanting to undo something one has worked hard on). What precise facet of this natural language behavior is the primary driver behind any such inferences still remains to be investigated, but regardless the evidence supports the hypothesis that observed behavior can have a stronger effect on ascribed agency/patency than appearance in certain contexts.

To some extent, we do not find these results terribly surprising. Linguistic communication is a hallmark of human intelligence, and engaging in a natural language interaction with a robotic agent is strong evidence for some degree of cognitive capacity. Appearance may set the expectations for an interaction, but the behaviors displayed in the interaction itself determine whether or not those expectations are met, unmet, or exceeded.

2.2 The Consequences of Overselling

Given the view that the appearance of a robotic agent sets the expectation for an interaction, how grave are the consequences of falling to meet these expectations? I would contend that these consequences are fairly short-term and mild in comparison to the more serious dilemmas raised in [5]. A quite human-looking robotic agent that is socially inept and incapable of performing many tasks that are easy for a human may perhaps be profoundly disappointing. But this disappointment is less harmful than a long-term relationship a human might form with a more socially capable robot, in which much time, energy, and emotional investment is wasted by the human partner. Such a hypothetical relationship is built by a wide range of social *behaviors* that are exhibited over time (rather than appearance). As such, the consequences of overselling behavior, I would argue, may be worse than the consequences of overselling appearance. However, future research is needed to further clarify and refine these claims depending on the particular HRI context (i.e. different interaction scenarios and subject populations). To ensure future human-robot interactions result in ethically desirable consequences, both the effects of overselling appearance and behavior ought to be considered.

3. CONCLUSION

In this abstract, we have briefly presented the problem raised by people inferring capabilities (be they physical, cognitive, or social) in robotic agents that may not be present, and how this may lead to excessive ascriptions of moral agency and patency in these artificial agents, which can in turn lead to ethically undesirable consequences. Much attention to this point has been focused on potentially regulating robotic appearance to mitigate this problem. However, we have put forth the position that in certain contexts interactions observed behavior may play a stronger role in the inferences interactants make regarding robots than observed appearance. Note that this is not to say that the morphology

of robots cannot have ethically undesirable consequences. It is certainly possible that appearance can cause distress in humans (e.g. the Uncanny Valley effect [3]). However, this is a separate issue from ensuring that people do not ascribe unwarranted degrees of agency and patency to robots. Any prospective ethical guidelines or codes of conduct for researchers and developers in robotics and HRI should also be focused on not overselling robot behavior *in addition* to robot morphology.

4. ACKNOWLEDGMENTS

I would like to thank Matthias Scheutz for providing helpful feedback on this abstract.

5. REFERENCES

- [1] G. Briggs, B. Gessell, M. Dunlap, and M. Scheutz. Actions speak louder than looks: Does robot appearance affect human reactions to robot protest and distress? In *Robot and Human Interactive Communication, 2014 RO-MAN: The 23rd IEEE International Symposium on*, pages 1122–1127. IEEE, 2014.
- [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*, 6(3):343–355, 2014.
- [3] M. Mori, K. F. MacDorman, and N. Kageki. The uncanny valley [from the field]. *Robotics & Automation Magazine, IEEE*, 19(2):98–100, 2012.
- [4] L. D. Riek, T.-C. Rabinowitch, B. Chakrabarti, and P. Robinson. Empathizing with robots: Fellow feeling along the anthropomorphic spectrum. In *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on*, pages 1–6. IEEE, 2009.
- [5] M. Scheutz. The inherent dangers of unidirectional emotional bonds between humans and social robots. In P. Lin, G. Bekey, and K. Abney, editors, *Anthology on Robo-Ethics*. MIT Press, 2012.