From HHI to HRI: Which Facets of Ethical Decision-Making Should Inform a Robot?

Jason Borenstein, Georgia Institute of Technology (United States)
Arthur Melo Cruz, Penn State University (United States)
Alan Wagner, Penn State University (United States)
Ronald Arkin, Georgia Institute of Technology (United States)
International Conference on Computer Ethics: Philosophical Enquiry (CEPE) 2023, Chicago, IL

Keywords: Ethical decision-making, human-human interaction, human-robot interaction, robot ethics

Extended Abstract

Robots, humanoid and otherwise, are being created with the underlying motivation in many cases that they will either replace or complement activities performed by humans. It has been many years since robots were starting to be designed to take over "dull, dirty, or dangerous" tasks (e.g., Singer 2009). Over time, roboticists and others within computing communities have extended their ambitions to create technology that seeks to emulate more complex ranges of human-like behavior, potentially including the ability to participate in complicated conversations. Regardless of how sophisticated its functionality is, a robot should arguably be encoded with ethical decision-making parameters, especially if it is going to interact with or could potentially endanger a human being. Yet of course determining the nature and specification of such parameters raises many longstanding and difficult philosophical questions.

Within this context, our research team is investigating some of the potential parameters that could inform a robot's ethical decision-making processes when it interacts with humans. Roboticists could pursue many different design approaches in terms of constructing a robot's ethical decision-making architecture. Furthermore, what determines whether a behavior is "ethical" should presumably be influenced by the user's characteristics and various contextual factors. During our project, we examined a subset of user characteristics and contextual factors that may have a bearing on identifying what the ethical course of action is for a human and perhaps for a robot as well.

To provide additional background, our team has been undertaking an NSF grant funded research project that in part has been examining what survey participants indicate is the appropriate way to act in a small collection of human-human interaction (HHI) scenarios. There are two main groups of research participants that we have surveyed: (1) American adults and (2) ethics experts. Responses from the former group we refer to as "folk morality" while the latter is referred to "expert morality". A key aim of acquiring such information is to guide the design of a robot's ethical architecture that could be applied to human-robot interaction (HRI) scenarios (Chen et al. 2022a; Surendran et al. 2022).

The main scenarios that both folk and expert participants responded to during the first round of surveys are the appropriateness of allowing deception to occur: first, when playing a board game with a child and second, when teaching an older adult how to organize pills in a sorting container. During a later stage of the project, we administered an additional survey to a new cohort of folk participants; the more

recent version of the "folk" survey included the original scenarios along with a variation of the game playing scenario with an adult instead of a child and a new scenario that inquiries about the appropriateness of using deception when teaching a young child how to swim (Surendran et al. 2022).

There are many characteristics about the person with whom one interacts along with contextual factors embedded in the interaction that could arguably be relevant to ascertaining what an ethically appropriate course of action is, including in our scenarios which focus on the circumstances within which deception might be justifiable. One of the characteristics that might be relevant in terms of deciding what is ethical is the age of the person with whom one interacts. We sought to identify, and to some degree isolate, the importance of this characteristic through the variations in our scenarios. For example, in the second folk survey, we sought to assess whether the age of the person playing the board game (child or older adult) was a relevant factor when deciding whether deception is acceptable. More generally, this detail about a user (i.e., their age) could be ethically relevant given considerations such as autonomy; in other words, presumably an adult is entitled to respect for autonomy whereas that ethical principle may be less or not applicable when interacting with a child.

The emotional state of the user during an interaction is another facet that we sought to explore during the project. Within the set of survey questions about the game playing and pill sorting scenarios, a subset of the questions only varied in terms of whether the hypothetical person in question was calm or frustrated. The intent was to explore whether survey participants viewed such variations in emotional state as being a decisive factor in weighing the ethically appropriate course of action (Chen et al. 2022b). If so, this finding could potentially be carried over to the realm of HRI and inform the design of a robot's decision-making.

The perceived level of risk associated with each possible decision option is another important and ethically relevant dimension we sought to examine. For example, survey participants were presumably envisioning what the potential consequences are to the person who loses a board game versus winning it. Survey participants were also implicitly asked to anticipate what may transpire if someone fails to learn how to sort pills correctly. Our study is seeking to weigh the importance of this dimension as compared to others we examined such as age (Chen et al. 2022b).

An additional feature of the game playing scenario, that was not present in the pill sorting case, was whether the deceptive act is performed by the person in question (e.g., player one deliberately tries to lose the game) or whether that person allows someone else to perform the act (e.g., player one lets the child cheat to win and does not hold the child accountable for cheating). Arguably, helping a child to win by playing badly is less ethically egregious than allowing cheating to occur. Moreover, there seems to be dimensions of each scenario, such as whether it pertains to healthcare versus game playing, that are tied to which type of ethical framework might be applied to guide decision-making (Surendran et al. 2022).

An overarching thread derived from the previously mentioned considerations is seeking to identify

which factors might be relevant to informing and guiding a human's ethical decision-making process. Of course, our approach only focuses on a small number of such factors, and at this point, the issue, and associated data, has only been viewed through the lens of HHI. Whether insights derived from HHI carry over to the ethical appropriateness of HRI is largely an open question. But our goal here is to highlight some of the dimensions of ethical decision-making that warrant examination while the enterprise of encoding ethical robots proceeds.

References

Chen, S., Arkin, R. C., Borenstein, J., Wagner, A. R., & Cruz, A. M. 2022a. Case-Based Robotic Architecture
with Multiple Underlying Ethical Frameworks for Human-Robot Interaction. Proceedings of the Seventh International Conference on Robot Ethics and Standards (ICRES 2022).

Chen, S., Surendran, V., Wagner, A. R., Borenstein, J., & Arkin, R. C. 2022b. Toward Ethical Robotic Behavior in Human-Robot Interaction Scenarios. arXiv preprint arXiv:2206.10727.

Singer, Peter. 2009. Wired for War: The Robotics Revolution and Conflict in the 21st Century. Penguin Press.

Surendran, V., Cruz, A. M., Wagner, A. R., Borenstein J., Arkin, R. C., & Chen, S. 2022. Informing a Robot Ethics Architecture Through Folk and Expert Morality. Proceedings of the Seventh International Conference on Robot Ethics and Standards (ICRES 2022).

Acknowledgments

This material was supported by National Science Foundation grants No. 1849068 and 1848974. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.