



This is a repository copy of *Is honesty the best policy for mismatched partners? Aligning multi-modal affordances of a social robot: an opinion paper.*

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/191263/>

Version: Published Version

Article:

Huang, G. and Moore, R.K. orcid.org/0000-0003-0065-3311 (2022) Is honesty the best policy for mismatched partners? Aligning multi-modal affordances of a social robot: an opinion paper. *Frontiers in Virtual Reality*. ISSN 2673-4192

<https://doi.org/10.3389/frvir.2022.1020169>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>



OPEN ACCESS

EDITED BY
Evelien Heyselaar,
Radboud University, Netherlands

REVIEWED BY
Emmanuele Tidoni,
University of Hull, United Kingdom

*CORRESPONDENCE
Guanyu Huang,
ghuang10@sheffield.ac.uk

SPECIALTY SECTION
This article was submitted to Virtual Reality and Human Behaviour, a section of the journal Frontiers in Virtual Reality

RECEIVED 15 August 2022
ACCEPTED 01 September 2022
PUBLISHED 16 September 2022

CITATION
Huang G and Moore RK (2022), Is honesty the best policy for mismatched partners? Aligning multi-modal affordances of a social robot: An opinion paper. *Front. Virtual Real.* 3:1020169. doi: 10.3389/frvir.2022.1020169

COPYRIGHT
© 2022 Huang and Moore. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Is honesty the best policy for mismatched partners? Aligning multi-modal affordances of a social robot: An opinion paper

Guanyu Huang* and Roger K. Moore

Department of Computer Science, University of Sheffield, Sheffield, United Kingdom

KEYWORDS

social robot, affordance design, honest signals, use cases, internal capabilities

Introduction

Spoken interactions between a human user and an artificial device (such as a social robot) have attracted much attention in recent decades (Lison and Meena, 2014; Oracle, 2020). Shifting from automation robots in the industrial domain, social robots are expected to be used in social domains, such as the service industry, education, healthcare, and entertainment (Bartneck et al., 2020, p.163).

According to Darling (2016)'s definition, a social robot is "a physically embodied, autonomous agent that communicates and interacts with humans on an emotional level". Many features play important roles in interactions with a social robot, such as people's experience with technology products, expectations of social robots, interactional environments and other features such as a social robot's appearance, voice and behaviours. In this last regard, affordance design affects how people perceive a social robot and how such perception affects their behaviours and experiences. The term "affordance" was invented by ecological psychologist Gibson (1977), who proposed that our perception of what it is possible to do with objects is shaped by their form. Affordance indicates what users see and can do with an object in a given situation; it is about perceptual action possibilities in an environment (Matei, 2020).

A strong tendency in social robot affordance design is to make human-robot interaction (HRI) resemble human-human interaction (HHI). It is hoped in many studies that robots designed with anthropomorphic appearances and human-like cognitive behaviours can enable humans to interact with them in similar ways as they would interact with other humans, even to develop social bonds (Leite et al., 2013; Kahn et al., 2015; Koyama et al., 2017; Ligthart et al., 2018). However, there are concerns about this approach. In fact, speech-based artificial agents' conversational interaction with human users is far from natural, and the language used tends to be formulaic (Moore et al., 2016).

One of the reasons behind this is a significant change in the applications of spoken human-agent interaction (HAI) along the evolution of spoken language technology applications (Moore, 2017a). Compared with "command and control systems" of the 1970s and contemporary smartphone-based "personal assistants", social robots are expected to be used in more dynamic and open environments. This implies that users' expectations, demands and ways to interact with spoken agents differ

depending on the use case. What has succeeded before in real-time spoken HAI (e.g., voice command for specific uses) may not work well for social robots in some contexts. Additionally, a social robot's human-like affordances could be seen as "dishonest" because such signals hide the fact that a social robot has limited interactive capabilities and is a "mismatched" conversational partner (Moore, 2015; 2017b). What's more, the approach to constructing a robot by integrating off-the-shelf human-like technologies lacks an appreciation of the function and behaviour of speech in a broader theoretical framework (Moore, 2015).

This paper takes a step back to consider what human users look for when speaking to a social robot. It starts by looking at the nature and the process of spoken interactions. It then discusses why honesty is the best policy for a social robot in HRI. Furthermore, the arguments presented here support the hypothesis that aligning a social robot's external affordances coherently with internal capabilities can shape its usability and improve human users' experience in HRI.

Broader theoretical background of spoken interaction

What happens when we talk with each other?

Spoken interaction is a joint activity grounded in social needs to cooperate on a moment-by-moment basis (Holtgraves, 2013). In this joint activity, interlocutors need to solve so-called "dilemma of cooperation" (Smith, 2010). The dilemma refers to two problems. One is the commitment problem of ensuring other individuals' collaborative motivation is genuine. The other is the collaboration problem of coordinating each individual's efforts to complete the collective task. Here is where language plays a vital role in facilitating the resolution of these two problems (Smith, 2010). Language helps to solve the commitment problem by enabling participants to express, recognise and act on each other's intentions, ultimately leading to shared intentionality (Bratman, 1992; Searle, 1995; Tomasello and Carpenter, 2007). Language also helps to solve the collaboration problem by building up common ground, which is "the knowledge that the communicating parties both share and know they share" (Krauss and Fussell, 1990, p.112).

It is worth noting that this process is not linear. In Pierce and Corey (2009)'s review, the most dynamic model of conversations is the transactional model. In this model, interlocutors send and receive signals simultaneously. Additionally, their shared field of knowledge and experiences¹ could allow them to use less speech

or even a single sound to achieve a successful interaction result (Hawkins, 2003).

Why a social robot could be a mismatched conversational partner?

Based on the above literature, it is safe to say that "sociality" in spoken interactions means collaboration. The efforts to achieve an effective collaboration lie in a pre-conversation shared field of knowledge and experiences, communicative competencies to align information in the conversation and post-conversation long-term memories, which updates the shared field of knowledge and experiences. Thus, interlocutors with similar shared field of knowledge and experiences, communicative competencies and memory capabilities can be considered "matched partners". Otherwise, interlocutors can be seen as "mismatched partners", like first and second language speakers, parents and babies, and humans and animals.

Hence, looking back at the case of social robots and other spoken artificial agents, it becomes clear why they are mismatched partners in HRI. To start with, robot designers and engineers must build the hardware and software to equip robots with desired abilities. This poses questions for them. For example, what shared fields of knowledge and experiences does the robot need to have? What sensory data of users and interactive environment does it need to get, and how can it act on collected information? What and when should it talk? How to use multi-modal cues to deliver the same message?

Without appropriate answers to these questions, it would be difficult for ordinary users to know how to coordinate their efforts to achieve a successful interaction with a social robot. For example, when talking with a social robot, how do we know what to say or how to speak? How do we know whether and how to adjust our behaviours to suit the agent? How do we know whether or when to give up?

Affordance design and its consequences

Human-like: To be or not to be

Bearing the above questions in mind, it is proposed that the human-like design of a social robot could be problematic by the following three arguments.

First, human-like design can be deceptive because it uses anthropomorphic signals that violate the associated humans' expectations, and it is for ulterior purposes (Danaher, 2020). When interacting with a social robot with human-like affordances, people may instantly perceive such an agent as a matched conversational partner. However, it is not the case. For

¹ The shared field of knowledge and experiences is not necessarily known to be shared by communicators (Bangertner and Mayor, 2013), which makes it different from the concept of "common ground"

example, the authenticity of a social robot's expressions can be doubted (Bartneck et al., 2020, p.195). Hence, it can be said that human-like cues risk overstating a social robot's capabilities. It then leaves people with negative impressions of the robot, resulting in deflated motivation to interact with it. One example is Ham and Midden (2014)'s study about people's negative reaction to a robot's deceptive praising. Hence, a social robot with such a misleading design can be seen as "dishonest", which is a primary ethical concern (Elder, 2016; Leong and Selinger, 2019; Hildt, 2021).

Second, human-like cues do not necessarily contribute to human users' tendency to anthropomorphise. Anthropomorphisation is a natural outgrowth of humans' social interaction and cognition (Bartneck et al., 2020, p.48). Long before social robots, evidence shows that people tend to humanise nonhuman entities regardless of their forms. For example, humans attribute mental states to animated geometrical shapes (Heider and Simmel, 1944), and people treat any technological form as social actors (Reeves and Nass, 1996). Hence, human users' tendency to anthropomorphise exists regardless of whether a social robot is human-like or not.

Third, current technologies are not advanced enough to deliver human-like perceptual cues concordantly. Social robots tend to be multi-dimensional. However, component technologies have not developed coherently. According to the studies of Moore (2012), Meah and Moore (2014) and MacDorman and Chattopadhyay (2016), the inconsistency of perceptual cues causes perceptual conflict, leads to uncertainty in HRI and contributes to the Uncanny Valley Effect (Mori, 1970). Also, it can cause human users to fall into the habitability gap where usability drops significantly when flexibility increases (Moore, 2017b). Therefore, human-like social robots have a high risk of failure.

Use of honest signals in HRI design: A hypothetical approach

Given the above analysis of should not and couldn't of human-like design, what is an appropriate way forward? One possibility is to explore a more appropriate affordance design for mismatched social agents. Here, it is hypothesised that the effectiveness of HRI could be improved by ensuring that a social robot's affordances are designed as "honest signals".

The concept of honest signals originally derived from evolutionary biology but has then developed in the social sciences (Pentland, 2010; Vinciarelli et al., 2011). Such signals refer to biological signals which cannot be faked thus convey reliable, useful information to the receiver; they are adopted as an evolutionarily stable strategy. In social communication, honest signals are unconsciously generated signals that

indicate signallers' genuine intentions or thoughts. They are reliable social signals because they are rooted in human brain structure, and biology (Pentland, 2010, p.3–4). By creating a direct or indirect link between perception and action, honest signals help make action possibilities be perceived appropriately so that signal receivers would have reasonable expectations and act accordingly. A successful example is the biomimetic robot MiRo's voice design, which aligns its physical and behavioural affordances (Moore and Mitchinson, 2017).

Hence, it is suggested that a true robot's affordances should be designed as honest signals. Such signals can then reflect a social robot's inner capabilities and help users form a dependable affordance prediction. By doing so, it can reduce communicative uncertainty and improve communicative effectiveness.

How honest is good enough?

Given that the principles for honest affordance design lie with the robot's internal capabilities, these must be quantifiable. It is worth noting that limited capabilities do not mean limited usability (Marge et al., 2020). Campa (2016) emphasised that two essential concepts for robots' usability are "scenario" and "persona". Both concepts are linked with user needs: one is what users need in given situations; the other is about expected user-robot relations and what users want from this relation. Hence, it is hypothesised that 1) similar user needs can be found in different domains; 2) the bottom line of a social robot's capabilities shall meet given user needs. Thus, a better scientific understanding of various use cases is needed (Marge et al., 2020) to help explore user needs, provide design guidance for a robot's capabilities and correlated affordance design, and develop a unified evaluation framework. Furthermore, it is proposed that honesty also exists in context, for example, Bonial et al. (2021)'s study about situated dialogue showed "*the surrounding physical context, as well as the dialogue history and some assumptions relevant to the robot's own embodied form and capabilities*" matter. Apart from that, another factor to consider is the role played by a spoken agent. Is it supposed to maintain the authenticity of a real person like a Holocaust survivor (Traum et al., 2015) or fictional characters (Gustafson et al., 1999; Leuski et al., 2006; Clark and Fischer, 2022)?

Finally, one thing to add is that the "honest level" for each component in a social robot may not be even due to technological or practical constraints. However, honest signals are helpful to reduce uncertainty if they are correct "on average" (Johnstone, 1999). Hence, a key question is how to measure a social robot's overall honesty and the honesty of its individual components.

Conclusion

What a social robot should look like, sound like and behave like is not only a practical concern, but also raises ethical issues in HRI. It is essential to understand what affects a social robot's affordance design and how affordance designs impact the effectiveness of HRI. This paper attempts to facilitate theoretical understanding by looking at the nature and process of spoken interaction and arguing that 1) a social robot is a mismatched partner when talking with human users; 2) a more appropriate affordance design for a social robot should be honest and aligned with its inner capabilities and states; 3) the honest design should be based on user needs in given use cases. It is argued that a social robot with honest affordances can explicitly represent its capabilities and states, shape users' expectations and reduce uncertainties during the interaction. While this helps us see the potential benefits of honesty in design, it also gives rise to plenty of challenges to overcome. For example.

- How could user needs and use cases be categorised more effectively?
- How do user needs affect people's expectations of a social robot's capabilities?
- How can a robot's capabilities be adjusted?
- How do a social robot's affordances reflect its capabilities?
- How would people perceive a social robot with honest affordance design in reality?
- Is it possible to develop an evaluation framework for social robot's affordances, capabilities and usability?

References

- Bangerter, A., and Mayor, E. (2013). "14 interactional theories of communication," in *Theories and models of communication. Handbook of communication science*, 257–271.
- Bartneck, C., Belpaeme, T., Eyssel, F., Kanda, T., Keijsers, M., and Šabanović, S. (2020). *Human-robot interaction: An introduction*. Cambridge: Cambridge University Press.
- Bonial, C., Abrams, M., Baker, A. L., Hudson, T., Lukin, S. M., Traum, D., et al. (2021). "Context is key: Annotating situated dialogue relations in multi-floor dialogue," in Proceedings of the 25th Workshop on the Semantics and Pragmatics of Dialogue.
- Bratman, M. E. (1992). Shared cooperative activity. *philosophical Rev.* 101, 327–341. doi:10.2307/2185537
- Campa, R. (2016). The rise of social robots: A review of the recent literature. *J. Eth. Emerg. Tech.* 26, 106–113. doi:10.55613/jeet.v26i1.55
- Clark, H. H., and Fischer, K. (2022). Social robots as depictions of social agents. *Behav. Brain Sci.* 2022, 1–33. doi:10.1017/s0140525x22000668
- Danaher, J. (2020). Robot betrayal: A guide to the ethics of robotic deception. *Ethics Inf. Technol.* 22, 117–128. doi:10.1007/s10676-019-09520-3
- Darling, K. (2016). "Extending legal protection to social robots: The effects of anthropomorphism, empathy, and violent behavior towards robotic objects," in *Robot law* (Cheltenham, UK: Edward Elgar Publishing).
- Elder, A. (2016). False friends and false coinage: A tool for navigating the ethics of sociable robots. *SIGCAS Comput. Soc.* 45, 248–254. doi:10.1145/2874239.2874274
- Gibson, J. J. (1977). Himba epochs. *Hist. Afr.* 1, 67–121. doi:10.2307/3171580
- Gustafson, J., Lindberg, N., and Lundeberg, M. (1999). "The august spoken dialogue system," in Sixth European Conference on Speech Communication and Technology (Citeseer).
- Ham, J., and Midden, C. J. (2014). A persuasive robot to stimulate energy conservation: The influence of positive and negative social feedback and task similarity on energy-consumption behavior. *Int. J. Soc. Robot.* 6, 163–171. doi:10.1007/s12369-013-0205-z
- Hawkins, S. (2003). Roles and representations of systematic fine phonetic detail in speech understanding. *J. phonetics* 31, 373–405. doi:10.1016/j.wocn.2003.09.006
- Heider, F., and Simmel, M. (1944). An experimental study of apparent behavior. *Am. J. Psychol.* 57, 243–259. doi:10.2307/1416950
- Hildt, E. (2021). What sort of robots do we want to interact with? Reflecting on the human side of human-artificial intelligence interaction. *Front. Comput. Sci.* 3. doi:10.3389/fcomp.2021.671012
- Holtgraves, T. M. (2013). *Language as social action: Social psychology and language use*. London, England, UK: Psychology Press.
- Johnstone, R. A. (1999). Signaling of need, sibling competition, and the cost of honesty. *Proc. Natl. Acad. Sci. U. S. A.* 96, 12644–12649. doi:10.1073/pnas.96.22.12644
- Kahn, P. H., Jr, Kanda, T., Ishiguro, H., Gill, B. T., Shen, S., Gary, H. E., et al. (2015). "Will people keep the secret of a humanoid robot? Psychological intimacy in hri," in Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction, 173–180.
- Koyama, N., Tanaka, K., Ogawa, K., and Ishiguro, H. (2017). "Emotional or social? How to enhance human-robot social bonding," in Proceedings of the 5th International Conference on Human Agent Interaction, 203–211.

Author contributions

GH and RM: article conceptualisation. GH: writing. RM: review. Both authors contributed to the article and approved the submitted version.

Funding

This work is supported by the Centre for Doctoral Training in Speech and Language Technologies (SLT) and their Applications funded by UK Research and Innovation [grant number EP/S023062/1].

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Krauss, R. M., and Fussell, S. R. (1990). "Mutual knowledge and communicative effectiveness," in *Intellectual teamwork: Social and technological foundations of cooperative work*, 111–146.
- Leite, I., Pereira, A., Mascarenhas, S., Martinho, C., Prada, R., and Paiva, A. (2013). The influence of empathy in human–robot relations. *Int. J. human-computer Stud.* 71, 250–260. doi:10.1016/j.ijhcs.2012.09.005
- Leong, B., and Selinger, E. (2019). "Robot eyes wide shut: Understanding dishonest anthropomorphism," in Proceedings of the conference on fairness, accountability, and transparency, 299–308.
- Leuski, A., Pair, J., Traum, D., McNerney, P. J., Georgiou, P., and Patel, R. (2006). "How to talk to a hologram," in IUI '06: Proceedings of the 11th International Conference on Intelligent User Interfaces (New York, NY, USA: Association for Computing Machinery), 360–362. doi:10.1145/1111449.1111537
- Lighthart, M., Hindriks, K., and Neerincx, M. A. (2018). "Reducing stress by bonding with a social robot: Towards autonomous long-term child-robot interaction," in Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction, 305–306.
- Lison, P., and Meena, R. (2014). Spoken dialogue systems: The new frontier in human-computer interaction. *XRDS Crossroads, ACM Mag. Students* 21, 46–51. doi:10.1145/2659891
- MacDorman, K. F., and Chattopadhyay, D. (2016). Reducing consistency in human realism increases the uncanny valley effect; increasing category uncertainty does not. *Cognition* 146, 190–205. doi:10.1016/j.cognition.2015.09.019
- Marge, M., Espy-Wilson, C., and Ward, N. (2020). *Spoken language interaction with robots: Research issues and recommendations, report from the nsf future directions workshop*. arXiv preprint arXiv:2011.05533.
- Matei, S. A. (2020). *What is affordance theory and how can it be used in communication research?* arXiv preprint arXiv:2003.02307.
- Meah, L. F., and Moore, R. K. (2014). "The Uncanny Valley: A focus on misaligned cues," in International Conference on Social Robotics (Berlin, Germany: Springer), 256–265.
- Moore, R. K. (2012). A Bayesian explanation of the 'uncanny valley' effect and related psychological phenomena. *Sci. Rep.* 2, 864–865. doi:10.1038/srep00864
- Moore, R. K. (2017a). "Appropriate voices for artefacts: Some key insights," in 1st International workshop on vocal interactivity in-and-between humans, animals and robots.
- Moore, R. K. (2015). From talking and listening robots to intelligent communicative machines. *Robots that talk listen* 2015, 317–335.
- Moore, R. K. (2017b). "Is spoken language all-or-nothing? Implications for future speech-based human-machine interaction," in *Dialogues with social robots* (Berlin, Germany: Springer), 281–291.
- Moore, R. K., Li, H., and Liao, S. H. (2016). "Progress and prospects for spoken language technology: What ordinary people think," in INTERSPEECH, San Francisco, CA, 3007–3011.
- Moore, R. K., and Mitchinson, B. (2017). "A biomimetic vocalisation system for miro," in Conference on biomimetic and biohybrid systems (Berlin, Germany: Springer), 363–374.
- Mori, M. (1970). Bukimi no tani [the uncanny valley]. *Energy* 7, 33–35.
- [Dataset] Oracle, W. (2020). *As uncertainty remains, anxiety and stress reach a tipping point at work*.
- Pentland, A. (2010). *Honest signals: How they shape our world*. Cambridge, MA, USA: MIT press.
- Pierce, T., and Corey, A. M. (2009). *The evolution of human communication: From theory to practice*. St. Louis, Missouri, USA: EtrePress.
- Reeves, B., and Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people*. Cambridge, UK: Cambridge University Press, 236605.
- Searle, J. R. (1995). *The construction of social reality*. New York, NY: Free Press.
- Smith, E. A. (2010). Communication and collective action: Language and the evolution of human cooperation. *Evol. Hum. Behav.* 31, 231–245. doi:10.1016/j.evolhumbehav.2010.03.001
- Tomasello, M., and Carpenter, M. (2007). Shared intentionality. *Dev. Sci.* 10, 121–125. doi:10.1111/j.1467-7687.2007.00573.x
- Traum, D., Jones, A., Hays, K., Maio, H., Alexander, O., Artstein, R., et al. (2015). "New dimensions in testimony: Digitally preserving a holocaust survivor's interactive storytelling," in International Conference on Interactive Digital Storytelling (Berlin, Germany: Springer), 269–281.
- Vinciarelli, A., Pantic, M., Heylen, D., Pelachaud, C., Poggi, I., D'Errico, F., et al. (2011). Bridging the gap between social animal and unsocial machine: A survey of social signal processing. *IEEE Trans. Affect. Comput.* 3, 69–87. doi:10.1109/t-affc.2011.27