ICRES 2019: International Conference on Robot
Ethics and Standards, London, UK, 29-30 July 2019.
https://doi.org/10.13180/icres.2019.29-30.07.017

SOCIAL INFLUENCE AND DECEPTION IN SOCIALLY ASSISTIVE ROBOTICS

K. WINKLE and A. VAN MARIS

Bristol Robotics Laboratory, University of the West of England
Bristol, United Kingdom Email: k.winkle@bristol.ac.uk

Keywords: Social Influence; Deception; Socially Assistive Robotics (SARs)

1. Robot Social Influence: Necessary but Deceptive?

SARs provide assistance through their social interaction;[1] typically through prompting and/or encouraging particular user behaviour(s). We suggest this is leveraging of the robot's social influence, and that many works in social/socially assistive robotics can be considered attempts to alter how a robot is perceived, such that it has greater social influence. This perception manipulation might be at odds with the latest ethical robot guidelines due to the inherent potential for deception with regards to the robot's capabilities.[2]

According to Shim & Arkin,[3] deception can be approached through different perspectives: philosophy, psychology, economics, military, cyberspace and biology. Both the perspective of philosophy and biology discuss the division of deception into either unintentional or intentional.[4] Intentional deception occurs when the deceiver is aware of the fact that a certain feature will raise false expectations. This is called behavioural deception, as it is often the behaviour from the deceived shows that causes the formation of these expectations. Unintentional deception occurs when a certain feature of the (unintentional) deceiver causes expectations that the deceiver means to evoke. This is also known as physical deception. Attempts to generate and leverage social influence could come under either of these categories, and can similarly be intentional or not. Therefore, we argue for a framework which establishes several levels of deception, and to what extent they are acceptable. In a first step towards this, we present a simple taxonomy of approaches as currently evidenced in the SAR literature.

2. A Taxonomy of Robot Social Influence & Deception

We have identified three key approaches regarding the generation and use of robot social influence, varying in intent to deceive and to generate social influence:

- **Natural Interaction:** Use of basic human-inspired communication cues such as gaze and gesturing. The motivation for employing such behaviours is typically to facilitate natural, effective communication rather than increasing social influence. Such cues could lead to some over-estimation of robot capabilities, but do not directly suggest missing (e.g. social, emotional) capabilities.
- **Implicit Anthropomorphism:** Purposeful design of behaviours which implicitly invoke anthropomorphism. Such behaviours are generally implemented via the natural interaction cues as described above, but go further in that are designed to have some 'meaning' - e.g. to indicate some robot 'thought' or 'feeling', e.g. modulating non-verbal and movement parameters to suggest 'personality' or portray emotion.

- **Direct Manipulation:** Behaviours (including dialogue) explicitly designed to make the robot appear more autonomous/intelligent, or to have (social, emotional) capabilities which it doesn't and/or behaviours which explicitly leverage social influence through e.g. through emotional appeals or relationship building/rapport such that the user's behaviour appears to have direct consequences for e.g. the robot's 'feelings'.

In summary, we stress the need for a framework that recognises variation in the intention both to deceive and manipulate with regards to robot social influence. Such a framework would prove a useful tool in the proper consideration of ethical risk associated with SARs.

References

1. D. Feil-Seifer and M. J. Mataric, Defining socially assistive robotics, in *9th International Conference on Rehabilitation Robotics, 2005. ICORR 2005.*, 2005.
2. BSI-2016 (2016).
3. J. Shim and R. C. Arkin, A taxonomy of robot deception and its benefits in hri, in *Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on*, 2013.
4. A. Dragan, R. Holladay and S. Srinivasa, *Autonomous Robots* **39**, 331 (2015).