
Proceedings Paper:
Cameron, D. orcid.org/0000-0001-8923-5591, Collins, E.C., Cheung, H. et al. (3 more authors) (2016) Don't Worry, We'll Get There: Developing Robot Personalities to Maintain User Interaction After Robot Error. In: Biomimetic and Biohybrid Systems. Living Machines, July 19-22, 2016, Edinburgh, UK. Lecture Notes in Computer Science, 9793 . Springer International Publishing , pp. 409-412. ISSN: 0302-9743 EISSN: UNSPECIFIED

https://doi.org/10.1007/978-3-319-42417-0_38

Don't worry, we'll get there: Developing robot personalities to maintain user interaction after robot error.

David Cameron, Emily Collins, Hugo Cheung, Adriel Chua,
Jonathan M. Aitken, and James Law

Sheffield Robotics, University of Sheffield, Sheffield, UK
{d.s.cameron, e.c.collins, mcheung3, dxachua1, jonathan.aitken, j.law}
@sheffield.ac.uk,
WWW home page: http://www.sheffieldrobotics.ac.uk/

Abstract. Human robot interaction (HRI) often considers the human impact of a robot serving to assist a human in achieving their goal or a shared task. There are many circumstances though during HRI in which a robot may make errors that are inconvenient or even detrimental to human partners. Using the ROBOtic GUidance and Interaction DEvelopment (ROBO-GUIDE) model on the Pioneer LX platform as a case study, and insights from social psychology, we examine key factors for a robot that has made such a mistake, ensuring preservation of individuals' perceived competence of the robot, and individuals' trust towards the robot. We outline an experimental approach to test these proposals.

Keywords: human-robot interaction, design, guidance, psychology

1 Background

Human-robot interaction (HRI) research typically explores interactions in which the robot plays a supportive or collaborative role for the human user [4]. However, there are circumstances in which robots may fail to meet these requirements, either through errors in processing the interaction scenario, or failure to adapt to changing HRI scenario circumstances. Furthermore, reliability and error rates of robots have both been identified as important factors in user trust towards robots [4]. Recent work has explored the social impact of a robot's fault or error, in terms of user cooperation [9], and whether an apology from a robot can mitigate the negative impact of the mistake [10]. However, there still remains much to be explored: first in terms of the negative impact even simple robot mistakes can have on user trust and their willingness to engage in HRI; and second, if the methods by which a robot acknowledges a mistake, and then potentially corrects for it, have differential impacts on HRI.

Recent work identifies that the means by which a socially adaptive robot asks for help can impact on: users' attitudes towards the robot, clarity in the support the robot needs, and people's willingness to use the robot in collaborative

tasks [2]. The context for that interaction does not concern robot error but rather robots requiring user intervention (completing a task outside of the robot's capability) to progress towards a goal. Nevertheless, the principles on which that interaction is based can be drawn upon to identify means by which robots can use particular social interactions to recover, in part, from mistakes.

The synthetic personality a robot exhibits can have substantial impact on the user's experience of, and their engagement with, HRI [3]. The relatively new social domain of HRI is unfamiliar to many, but principles of social psychology have been applied to social HRI scenarios with promising results. These include: accurate recognition of synthetic personality types, even in non-humanoid robots [6]; development of classification of social 'rules' for robots to adhere to [3]; and participant response towards synthetic personalities corresponding with theoretical models of interpersonal cooperation [2]. This paper draws on social psychological theories to develop a model of social factors for robots that support recovery from robot error and maintain user interaction.

2 ROBO-GUIDE interactive scenario

To explore the social factors that support a robot in recovery from error, and maintain user interaction, it is useful to consider an interactive scenario in which these circumstances might arise. The ROBOtic GUidance and Interaction DEvelopment (ROBO-GUIDE) project [1, 8] is an ideal scenario to consider the impact of such social factors as it requires humans to place trust in a robot, even in the instance where an error might occur.

ROBO-GUIDE is embodied on the Pioneer LX mobile platform. The platform is capable of autonomously navigating a multistory building and leading users from their arrival point to their desired destination. Our focus here is a critical point in building navigation: floor determination whilst using the elevator to navigate between floors. Each floor in the building is similar and ROBO-GUIDE relies on subtle cues to differentiate between them; in noisy environments errors can occur [8]. For example, during busy periods the corridor structural features or floor-indication signs may be obscured, noise may mask lift announcements, or the ROBO-GUIDE might be misinformed or misled by a member of the public.

We identify the disembarkation of the elevator on the wrong floor as a simple and, critically, natural type of error for users to encounter (for the purposes of the experiment, errors would be staged). This tour guide scenario, in which individuals use a robot to navigate an unfamiliar building, means any error would mildly inconvenience the user. Moreover, it gives opportunity for the robot to recover from the error and allows testing of different means for the robot to socially communicate error and attempts to correct it.

3 Social-recovery and experimental proposal

Previous HRI work has identified a social psychological model of cooperation [7] as a useful framework for exploring the impact different personalities can

have on user willingness to engage with robots [1]. This model, when applied in an HRI context, contrasts the impact of friendly-oriented statements to build user-liking, and goal-oriented statements to suggest the robot's task competency. Findings indicate that individuals are more willing to use a robot they like than a robot suggesting task competency; they also regard the interaction with a friendly robot to be less ambiguous [2]. Individuals further report trust towards the robot across the dimensions offered in the model: affective (from personable interactions) and cognitive (from evidence of competency). While results do not show substantive differences in trust between conditions, this may be due to a ceiling effect because there was no challenge made to the competency of the robot.

The error scenario (section 2) provides such a challenge to the robot's apparent competency; it further provides opportunity for the robot to attempt to socially recover from the error and work to restore use perceptions of competency. Again, using the framework developed in [1], friendly-oriented and goal-oriented synthetic personalities, as means to socially communicate error, may result in differences in individuals' views towards the robot.

The framework may be further enhanced by considering social psychological understanding on the impact of acknowledging and apologising for error. Apologies for errors in competency are observed to raise an individual's trustworthiness but not their apparent competency [5]. Simple apologies may therefore support a user's affective trust towards a robot but not their cognitive trust. In contrast, identification of the error and communication of means to resolve it, to maintain progress towards a goal, may restore users cognitive trust.

We outline a brief experimental proposal to test the impact of statements promoting affective and cognitive trust for the user in the error scenario (section 2). Acknowledgments of the error by the robot are planned to be manipulated in a 2x2 experimental between-subjects design: inclusion or absence of the competency-oriented apology-oriented statements following error. These will be communicated using the on-board speech synthesizer[1]. We anticipate that these statements will impact on participant willingness to use the robot through the key channels of affective and cognitive trust.

Participants will experience one of the four conditions: 1) the control condition comprising simple instructions for the user to follow after making an error, (e.g., 'Follow me back to the lift'); 2) inclusion of competency-oriented statements that emphasise the robot's ability to recognise the environment, identifying the cues used to orient, and reaffirming the goal (e.g., 'That sign said we are on C floor and we need to go to B floor. Follow me back to the lift'); 3) inclusion of apology-oriented statements that emphasise attempts to relate to users but *do not* indicate competency (e.g., 'Sorry about the error; we can all make mistakes sometimes. Follow me back to the lift'); 4) inclusion of both the competency- and apology-oriented statements.

[1] The viable alternatives of pre-recorded spoken phrases or an on-screen display are acknowledged

Outcomes are measured using the prior measures of affective and cognitive trust, clarity of interaction and willingess to use the robot [2]. We anticipate that the apology and competency statements will support affective and cognitive trust respectively, although affective trust better predict willingness to use the robot in future. Findings from this work will support the development of a socially adaptive robots for HRI and further reveal the social models users draw upon when interacting with socially engaging robots.

Acknowledgments

This work was supported by European Union Seventh Framework Programme (FP7-ICT-2013-10) under grant agreement no. 611971.

References

1. Cameron, D., Collins, E.C., Chua, A., Fernando, S., McAree, O., Martinez-Hernandez, U., Aitken, J.M., Boorman, L., Law, J.: Help! I cant reach the buttons: facilitating helping behaviors towards robots. In S. P. Wilson, P. F. M. J. Verschure, A. Mura & T. J. Prescott (Eds.). Biomimetic and Biohybrid Systems, LNAI 9222, 354-358. Springer. (2015)
2. Cameron, D., Loh, E.J., Chua, A., Collins, E.C., Aitken, J.M., Law, J.: Robot-stated limitations but not intentions promote user assistance. In M. Salem, A. Weiss, P. Baxter, & K. Dautenhahn (Eds.), Proceedings of the 5th International symposium on New Frontiers in Human-Robot Interaction. AISB. (In Press)
3. Dautenhahn, K.: Socially intelligent robots: dimensions of humanrobot interaction. Philosophical Transactions of the Royal Society of London B: Biological Sciences 362, 679–704. (2007)
4. Hancock, P.A., Billings, D.R., Schaefer, K.E., Chen, J. De Visser, E., Parasuraman, R.: A meta-analysis of factors affecting trust in human-robot interaction. The Journal of the Human Factors and Ergonomics Society 53 517–527 (2011)
5. Kim, P., Ferrin, D., Cooper, C., Dirks, K.: Removing the Shadow of Suspicion: The Effects of Apology Versus Denial for Repairing Competence- Versus Integrity-Based Trust Violations. Journal of Applied Psychology, Vol 89, 104–118. (2004)
6. Lee, K.M., Peng, W., Jin, S.A. and Yan, C.: Can robots manifest personality?: An empirical test of personality recognition, social responses, and social presence in humanrobot interaction. Journal of communication, 56 754–772. (2006)
7. McAllister, D.J.: Affect-and cognition-based trust as foundations for interpersonal cooperation in organizations. Academy of Management Journal, 38 24–59. (1995)
8. McAree, O, Aitken, J.M., Boorman, L., Cameron, D., Chua, A., Collins, E.C., Fernando, S., Law, J., Martinez-Hernandez, U.: Floor Determination in the operation of a lift by a mobile guide robot. In Proceedings of the European Conference on Mobile Robots. (2015)
9. Salem, M., Lakatos, G., Amirabdollahian, F., Dautenhahn, K.: Would You Trust a (Faulty) Robot? Effects of Error, Task Type and Personality on Human-Robot Cooperation and Trust. In: Proceedings of the 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI 2015), Portland, Oregon, USA (2015)
10. Snijders, D.: Robots Recovery from Invading Personal Space 23rd Twente Student Conference on IT, Enschede, The Netherlands (2015)