I Blame You Anyways!

An Experimental Study on the Perceived Responsibility of a Social Robot's Pre-Programmed Behavior

Aike C. Horstmann & Nicole C. Krämer

Social Psychology: Media and Communication, University of Duisburg-Essen

Author Note

Nicole C. Krämer https://orcid.org/0000-0001-7535-870X

Aike C. Horstmann https://orcid.org/0000-0003-4693-1743

We have no conflict of interest to disclose.

Correspondence concerning this article should be addressed to Aike C. Horstmann, Social Psychology: Media and Communication, University of Duisburg-Essen, Forsthausweg 2, 47057 Duisburg, Germany. Email: aike.horstmann@web.de

DOI: 10.18154/RWTH-CONV-245959

I Blame You Anyways!

An Experimental Study on the Perceived Responsibility of a Social Robot's Pre-Programmed Behavior

Since social robots are rapidly advancing, they increasingly enter people's everyday environments such as homes, work places, schools, and hospitals (Broekens et al., 2009; Dautenhahn et al., 2005). Consequently, human-robot interactions are becoming more socially situated and multi-faceted (Dautenhahn, 2007; Young et al., 2011). With this increase in social interactions, the question arises how people perceive and evaluate the robots' actions, particularly regarding responsibility. According to attribution theory, people constantly attempt to understand the causes and implications of others' behaviors (Kelley, 1973). Different factors determine whether (problematic) behavior is attributed to an internal or external cause (Kelley, 1973). Against this background, circumstances influencing whether a robot or other factors are considered responsible for its behavior are examined.

Various studies show that people expect artificial entities to adhere to social norms such as politeness (Fogg & Nass, 1997; Nass et al., 1999). With a robot's behavior being pre-programmed, not the robot but the programmer(s) should be perceived responsible for impolite behavior since they pose a high external justification for the robot's behavior (Kelley, 1973). Here, we hypothesize that when a social robot's feedback is impolite and believed to be pre-programmed compared to autonomous, this leads to a more positive evaluation of the robot and the interaction with it.

In addition, people's feelings towards social robots are often mixed due to the prevalence of two contradicting prospects for them (Bruckenberger et al., 2013; Scopelliti et al., 2005). Many people fear that robots will become competitors to them, mainly professionally but also personally (Khan, 1998; Ray et al., 2008). In contrast, social robots are also expected to make life easier by functioning as helpful assistants (Horstmann & Krämer, 2019; Oestreicher & Eklundh, 2006). Since previous research showed that a robot is evaluated differently after being framed as threatening or helpful (Horstmann & Krämer, 2020), we hypothesize that when a social robot is believed to become an assistant compared to a competitor, the robot's impolite feedback leads to a more negative evaluation.

In an experimental online study with a 2x2x2 between-subjects-design (N = 394), people read a vignette describing the social robot Pepper either as future assistant (will be very helpful and assist humans with many exhausting tasks) or competitor (will take over many tasks currently executed by humans). Furthermore, they were told that they will receive feedback by Pepper that was either described to be generated autonomously by the robot itself or to be pre-programmed by programmers in advance. During a subsequent quiz, Pepper either provided positive (e.g., "I am impressed by how well you are doing!") or negative (e.g., "I am impressed by how bad you are at this.") feedback. Effects on the evaluation of the robot (its agency, responsibility, competence, and sociability) and the interaction with it are measured. These self-reports are extended by an implicit measure of the automatic activation of negative or positive attitudes towards Pepper and robots in general. The study will deliver valuable insights to successfully design human-robot interactions by avoiding and diminishing misunderstandings and subsequent frustrations.

References

Broekens, J., Heerink, M., & Rosendal, H. (2009). Assistive social robots in elderly care: a review. *Gerontechnology*, *8*(2), 94–103. https://doi.org/10.4017/gt.2009.08.02.002.00

Bruckenberger, U., Weiss, A., Mirnig, N., Strasser, E., Stadler, S., & Tscheligi, M. (2013). The good, the bad, the weird: Audience evaluation of a "real" robot in relation to science fiction and mass media. In M. Jamshidi (Ed.), *Advance Trends in Soft Computing: Proceedings of the World Conference on Soft Computing - WCSC '13* (Vol. 8239, pp. 301–310). Springer. https://doi.org/10.1007/978-3-319-02675-6_30

Dautenhahn, K. (2007). Socially intelligent robots: Dimensions of human-robot interaction. *Philosophical Transactions of the Royal Society of London*, *362*(1480), 679–704. https://doi.org/10.1098/rstb.2006.2004

Dautenhahn, K., Woods, S. N., Kaouri, C., Walters, M. L., Koay, K. L., & Werry, I. (2005). What is a robot companion - Friend, assistant or butler? In *Proceedings of the 18th IEEE/RSJ International Conference on Intelligent Robots and Systems - IROS '05* (pp. 1192–1197). IEEE. https://doi.org/10.1109/IROS.2005.1545189

Fogg, B. J., & Nass, C. I. (1997). Silicon sycophants: the effects of computers that flatter. *International Journal of Human-Computer Studies*, *46*(5), 551–561. https://doi.org/10.1006/ijhc.1996.0104

Horstmann, A. C., & Krämer, N. C. (2019). Great expectations? Relation of previous experiences with social robots in real life or in the media and expectancies based

on qualitative and quantitative assessment. *Frontiers in Psychology*, *10*, 939. https://doi.org/10.3389/fpsyg.2019.00939

Horstmann, A. C., & Krämer, N. C. (2020). Expectations vs. Actual behavior of a social robot: An experimental investigation of the effects of a social robot's interaction skill level and its expected future role on people's evaluations. *PloS One*, *15*(8), e0238133. https://doi.org/10.1371/journal.pone.0238133

Kelley, H. H. (1973). The processes of causal attribution. *American Psychologist*, *28*(2), 107–128. https://doi.org/10.1037/h0034225

Khan, Z. (1998). *Attitudes towards intelligent service robots*. Royal Institute of Technology.

Nass, C. I., Moon, Y., & Carney, P. (1999). Are people polite to computers? Responses to computer-based interviewing systems. *Journal of Applied Social Psychology*, *29*(5), 1093–1109. https://doi.org/10.1111/j.1559-1816.1999.tb00142.x

Oestreicher, L., & Eklundh, K. (2006). User expectations on human-robot co-operation. In *Proceedings of the 15th IEEE International Symposium on Robot and Human Interactive Communication - RO-MAN '06* (pp. 91–96). IEEE. https://doi.org/10.1109/ROMAN.2006.314400

Ray, C., Mondada, F., & Siegwart, R. (2008). What do people expect from robots? In *Proceedings of the 21st IEEE/RSJ International Conference on Intelligent Robots and Systems - IROS '08* (pp. 3816–3821). IEEE. https://doi.org/10.1109/IROS.2008.4650714

Scopelliti, M., Giuliani, M. V., & Fornara, F. (2005). Robots in a domestic setting: a psychological approach. *Universal Access in the Information Society*, *4*(2), 146–155. https://doi.org/10.1007/s10209-005-0118-1

Young, J. E., Sung, J., Voida, A., Sharlin, E., Igarashi, T., Christensen, H. I., & Grinter, R. E. (2011). Evaluating human-robot interaction: Focusing on the holistic interaction experience. *International Journal of Social Robotics*, *3*(1), 53–67. https://doi.org/10.1007/s12369-010-0081-8