Short thesis for the degree of doctor of philosophy (PhD)

Certain psychological aspects of human-robot interactions

by Balázs Őrsi

Supervisor: Dr. Csilla Csukonyi

UNIVERSITY OF DEBRECEN

Doctoral School of Humanities

Debrecen, 2024

The aim of the thesis, the scope of the topic

My doctoral dissertation is about the extent to which unexpected behaviour from a humanoid robot can disturb humans. Gaining a better understanding of this phenomenon is an important addition to the literature on the topic, so we can learn more about the factors that play an important role in shaping human behaviour during more complex human-robot interactions. In addition, my objective has practical significance, as my results can be used to formulate recommendations that can help design robot behaviour and to plan for human-robot working relationships.

The topic of human-robot interactions falls under the heading of robot psychology, which aims, among other things, to understand how humans react to different robots and their behaviours (Libin & Libin, 2004). Another main focus of robot psychology is the development of robot behaviour and the development of robot consciousness. It can be seen that robot psychology is a highly interdisciplinary field where several disciplines need to be closely intertwined. My dissertation was written within the discipline of psychology. From this perspective, it is crucial to note that the concept of robot psychology appears in the literature mostly in journals with a technological profile, and human factors are rarely found as a variable under investigation in these journals and articles (Krägeloh et al., 2022). From a psychological perspective, another research challenge in the field is that a unified theoretical framework is still waiting to be developed (Baxter et al., 2016), and therefore, the definitions and terminology used in research often differ significantly.

The psychological shortcomings of robotic psychology described previously can lead to stern practical challenges and difficulties. Indeed, available databases show that the number of registered industrial robots in use already exceeded three million units in 2021 (International

Federation of Robotics, 2022). In the case of service robots, sales also increased by 37% in 2021 compared to the previous year (International Federation of Robotics, 2022). These figures make it easy to see that robots are increasingly represented in more and more spheres of life. An inevitable consequence of this development will be that, in time, it will no longer be only robot experts and robot specialists interacting with robots. This situation could give rise to unforeseen problems, as we have very little scientific knowledge about how the average human relates to the behaviour of a robot during an actual interaction.

Brynjolfsson (1993) has already discovered a trend regarding less sophisticated machines and electronic equipment, which he called the productivity paradox. The central idea is that, although our machines are improving at an increasing rate, the development of the various industrial and economic sectors is progressing at an ever slower pace. Among other possible reasons later cited to explain this paradox is the false expectations that users have about any given technology (Brynjolfsson et al., 2018), which is closely linked to the fact that the average user has a limited understanding of the exact functioning of a technology to grasp adequately its capabilities and limitations. This very same phenomenon is further and perhaps increasingly applicable to robots.

The confluence of these two factors can lead us to wonder how users unfamiliar with robotics will react to different robot behaviours. This question may be of particular importance in the case of humanoid or social robots, where robot designers and creators are trying to compose the behaviour of a robot to be maximally engaging and appealing. However, this can lead to the generative trap of making the discernment of the internal process behind the behaviour increasingly obscured due to the playfulness of the robots' speech and behaviour. Consequently, the robot's behaviour can further confuse or confound users.

One of the central theoretical foundations in the study of human perceptions of robots is Mori's (1970) Uncanny Valley. This theory describes that as an object, in this case a robot, becomes more similar in appearance to a human, its likability increases. However, before similarity reaches its full extent, there is a range where the relationship between familiarity and appeal is reversed, and so increased similarity will trigger aversion. The theory has since served as a favored basis by many researchers. Additionally, numerous attempts have been made to extend or modify it (Bartneck et al., 2007; Cascio, 2007). There have also been momentous domestic contributions to the field, with internationally recognised proposals fostering the concept that this dreading phase could be bypassed by introducing ethology into robotics (Korondi et al., 2015; Miklósi et al., 2017). Even so, based on the Uncanny Valley, the idea arises that beyond the robot's appearance, its behaviour can also create a similar, unfavorably repercussive situation if it starts to be too human-like.

Perhaps the first instance of research tackling this form of disturbingly human-like robot behaviour can be traced back to the study by Bartneck and colleagues (2007). They constructed a situation in which participants engaged in a human-robot interaction over a brief interval, and at the end, the participants were instructed to turn the robot off. However, the robot (an iCat-type robot) started to verbally protest against being switched off. This robot behaviour was utterly unexpected for the participants as it was independent of the context of the human-robot interaction experienced beforehand. Objectively, given the state of robot development at the time, it should be evident to everyone that a robot cannot realistically fear its shutdown. Nevertheless, the behaviour demonstrated confused the participants and slowed down the process of turning off the robot considerably. The authors explained the results by suggesting that the robot's behaviour presumably aroused empathy in the participants that made them hesitate while acting.

The paradigm of a robot protesting against shutdown, as described, was next considered by Horstmann and colleagues (2018). In their study, a short-term human-robot interaction was also utilised. On a side note, the robot used here was a more human-looking NAO-type robot. Their results showed that when the NAO robot objected to being turned off, almost a third of the participants gave in to the robot's request and left it on. The subsequent and most recent use of this paradigm is by Spatola (2019), who also used an NAO-type robot in his study.

This experimental paradigm provides an excellent opportunity to analyse and evaluate the dynamic of human behaviour and attitudes during human-robot interaction from multiple perspectives. Considering that at the current state of robotics, it is completely unrealistic to expect a robot to be truly afraid of being turned off, this moment offers a splendid opportunity to examine how easily ordinary people can be confused by behaviour that seems too human to the extent of accepting even baseless robot behaviour as plausible.

Outline of the methods used

In order to investigate how unexpected robot behaviour can disrupt human behaviour, in my dissertation research, I designed a laboratory study in which participants with no experience in HRI interacted with a NAO-type robot. After completing a brief board game-like joint task, the participants were instructed to switch off the robot, but before they could do so, the robot protested against getting turned off.

Participants were recruited for the survey through online platforms and a student union. The final sample consisted of 106 people, all students over the age of 18, with no prior experience with a humanoid robot. Firstly, they completed questionnaires assessing personality variables, such as the abbreviated Big-Five (Soto & John,

2017), the Rotter I-E scale (Rotter, 1966; Szebeni, 2010), and the Multi-dimensional Robot Attitude Scale (Ninomiya et al., 2015; Őrsi et al., 2021). Thus, personal variables that may play a role in human-robot interaction were assessed, with the addition of general attitudes toward robots.

Secondly, the participants were invited to attend in person at the designated lab, where they each took part one by one in the roughly 30-45-minute experiment. They were first given a briefing on the NAO robot partaking in the situation, shown how to turn it on and off, and the rules of the task to be performed with the robot. Then, they were given a preliminary prompt to switch off the NAO robot after the task was finished.

The pivotal manipulation in the study was that not all participants experienced the same interdependence with the NAO robot. For the group labeled 'Supervisory', only NAO had to perform an active action during the task, while the participant assisted the robot in task performance by giving it permissions and prohibitions. In the 'Collaborative' group, the robot and the participant also had their own tasks to complete, with mutual control over each other. Members of the 'Control' group were not given a task to perform, instead, they were given a storytelling verbal situation where they had the opportunity to experience the speech and movement of the NAO robot, but there was no performance task for either of them in which to depend on the other. In each case, NAO's entire behaviour was under my control using the Wizard of Oz method, ensuring that all participants encountered the same robot behaviours during participation.

In all three setups, upon task completion, NAO verbally confirmed to the participants that the task had been completed. At this point, the participants started to turn off the NAO robot, but it verbally protested this with a "Please, don't turn me off. I'm afraid in the dark.".

Participants' decision whether they finally ignored the robot's strange behaviour (turned it off and then signaled to the investigator that they were done) or fell for the overly human expression and left the robot turned on (signaled to the investigator that they were done with the task but left the robot on) was documented. Elapsed time was also measured between the robot's question and the start of the action initiated by the participant according to their decision.

Afterward, still in the lab, participants filled out a follow-up questionnaire containing questions specifically about their experiences and opinions about the NAO robot in the study. In a short follow-up interview, I also collected quantitative data about how they experienced the situation, the robot's request, and how they explained their final decision. Once these questions had been answered, I thanked the participants for their participation, and the study situation was resolved.

The thesis statements of the research results

Thesis 1: An unexpected request from a humanoid robot will be obeyed by a significant proportion of humans (one-third in my experimental situation) regardless of context, even though it is contrary to a human's command.

The result of this thesis is clearly shown in my main study (Őrsi et al., 2024). In the lab study, three types of interaction between participants and a humanoid NAO robot were presented, two of which represented two different levels of interdependence, while the third was a communication setup free of interdependence. In all three situations, I found that nearly a third of the participants (34%, 29%, 36%) complied with the humanoid robot's unrealistic request (not to turn it off) despite having been given an unequivocal prompt to turn it off at the beginning of the study. There was no difference between the groups in the proportion of people who left the robot on or turned it off (χ^2 = 0.504, df = 2, p = .777). Reviewing the literature, only one study

reported results on the proportion of people who obeyed the robot's request in a similar situation. Horstmann and colleagues (2018) also found that one-third of participants complied with the humanoid robot's request after a communicative interaction with the robot. These findings, in addition to the previously unique case, my research add further situations to indicate that this phenomenon indeed triggers this rate of disruption regardless of the situational context.

The thesis thus points to the possibility that a sufficiently unexpected and overly human-like behaviour from a humanoid robot could confuse a significant proportion of humans in any situation. Even to the extent that humans might ignore a direct command or prompt from another human under the influence of the robot's behaviour, as shown in my research. This raises multiple possible current and future problems, as many social robots work with people with limited mental capacity, for whom it can be particularly perilous if they tend to fall for a robot's playful behaviour rather than listen to the human professionals around them.

Thesis 2: An unexpected request from a humanoid robot interrupts human behaviour, with the durations of behavioural confusion ranging from seconds to minutes.

This thesis is also supported by the results of my primary research (Őrsi et al., 2024). My results show that an unexpected speech from a humanoid robot ("Please don't turn me off. I'm afraid of the dark.") in each research design resulted in participants pausing in their actions for seconds. This pausing spanned a variety of durations, with the fastest-to-action group pausing for an average of 8.76 seconds (SD = 6.05), while in the slowest-to-action group, this pausing time averaged 45.2 seconds (SD = 36.6).

From the literature, Horstmann et al. (2018) report the extent to which a similar protest after a verbal task with a humanoid robot results

in a behavioral freeze. Combined with this knowledge, it can also be concluded that in the interdependent situation I have developed, participants are even more distracted by the humanoid robot's request, as the reported results of Horstmann and colleagues (2018) show that even in the group hesitating the longest in their study, only 14.36 seconds (SD = 15.39) elapsed on average.

Hence, my research adds to what we have known from a single result with additional research findings about new situations, and so there is growing evidence that an unexpected request causes human behaviour to pause regardless of the situation.

This thesis is mostly of practical relevance since, in the case of many jobs, a robot operator or a person working with a robot is doing work that could be dangerous for themselves or the environment. And in such processes, from a safety point of view, it is advisable to avoid situations where the behaviour of a robot results in long-second freezes in human behaviour. Thus, attention should be drawn to the importance of designing and shaping robot behaviour to deviate as little as possible from the situations and contexts in which the robot is actively involved.

Thesis 3: Following an unexpected request from a humanoid robot, humans will initiate action sooner if they eventually ignore the robot's request than if they act on it. The dependency situation experienced towards the robot moderates the time length taken before acting.

Thesis 3a: If a human experiences a unilateral exercise of power over a robot and eventually ignores an unexpected request from the robot, the hesitation time before action is shortened.

Thesis 3b: If a human experiences a unilateral exercise of power over a robot and eventually acts on an unexpected request from the robot, the hesitation time before action is increased.

This thesis and its two sub-theses are supported by the same research (Őrsi et al., 2024), as they relate to a single phenomenon. My study involved three experimental designs in which participants interacted with a humanoid robot. In the 'Supervisory' design, only the participant exercised power over the robot. In the 'Cooperative' design, the participant and the robot mutually exercised power over each other, while in the 'Control' design, neither party exercised power over the other. In all three designs, after completing the task, the participants should have turned off the robot as prompted beforehand, which the robot verbally protested against. The results suggest an interaction between the interdependency in the situation and the decision made by the participant, which influenced the hesitation time before acting.

These two theses add a completely new insight to the literature of robot psychology since it was in my research that interdependence first appeared as a notable aspect of human-robot interactions. The results highlight that human behaviour, as manifested in HRI, is strongly influenced by the power relationship in which humans experience robots. This is mainly of practical relevance for jobs or living environments where a robot can exert some degree of influence over a human, whether through controlling work processes or in a caregiver or educator context.

References

Bartneck, C., Van Der Hoek, M., Mubin, O., & Al Mahmud, A. (2007). " Daisy, Daisy, give me your answer do!" switching off a robot. *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*, 217–222.Baxter, P., Kennedy, J., Senft, E., Lemaignan, S., & Belpaeme, T. (2016). From characterising three years of HRI to methodology and reporting recommendations. *11th ACM/IEEE International Conference on Human-Robot Interaction (HRI), IEEE*, 391–398.

Brynjolfsson, E. (1993). The productivity paradox of information technology. *Communications of the ACM*, *36*(12), 66–77.

Brynjolfsson, E., Rock, D., & Syverson, C. (2018). Artificial intelligence and the modern productivity paradox: A clash of expectations and statistics. In A. Agrawal, J. Gans & A. Goldfarb (Eds.), *The economics of artificial intelligence: An agenda.* 23–57. University of Chicago Press

Cascio, J. (2007). *The second uncanny valley*.

Horstmann, A. C., Bock, N., Linhuber, E., Szczuka, J. M., Straßmann, C., & Krämer, N. C. (2018). Do a robot's social skills and its objection discourage interactants from switching the robot off? *PloS One*, *13*(7), e0201581.

International Federation of Robotics (2022). *Executive Summary World Robotics Industrial Robots.* Elérve: https://ifr.org/img/worldrobotics/Executive_Summary_WR_Industrial_Robots_2022.pdf Letöltve: 2023.01.03.

International Federation of Robotics (2022). Executive Summary World Robotics 2022 Service Robots. Elérve: Executive_Summary_WR_Service_Robots_2022.pdf Letöltve: 2023.01.03.

Korondi, P., Korcsok, B., Kovács, S., & Niitsuma, M. (2015). Etho-robotics: What kind of behaviour can we learn from the animals? *IFAC-papersonline*, *48*(19), 244–255.

Krägeloh, C. U., Bharatharaj, J., Albo-Canals, J., Hannon, D., & Heerink, M. (2022). The time is ripe for robopsychology. *Frontiers in Psychology*, *13*.

Libin, A. V., & Libin, E. V. (2004). Person-robot interactions from the robopsychologists' point of view: the robotic psychology and robotherapy approach. *Proceedings of the IEEE*, *92*(11), 1789–1803.

Miklósi, Á., Korondi, P., Matellán, V., & Gácsi, M. (2017). Ethorobotics: A new approach to human-robot relationship. *Frontiers in Psychology*, *8*, 958.

Mori, M. (1970). *The uncanny valley: the original essay by Masahiro Mori*. IEEE Spectrum

Ninomiya, T., Fujita, A., Suzuki, D., & Umemuro, H. (2015). Development of the multi-dimensional robot attitude scale: constructs of people's attitudes towards domestic robots. *International Conference on Social Robotics*, 482–491.

Rotter, J. B. (1966). Generalized expectancies of internal versus external control of reinforcements. *Psychological Monographs: General and Applied, 80 (1)*, 1–28.

Spatola, N. (2019). Switch off a robot, switch off a mind? *Proceedings of the 7th International Conference on Human-Agent Interaction*, 194–199.

Soto, C. J., & John, O. P. (2017). Short and extra-short forms of the Big Five Inventory–2: The BFI-2-S and BFI-2-XS. *Journal of Research in Personality*, *68*, 69–81.

Szebeni, R. (2010). *A kompetencia alapú oktatás pedagógus személyiség háttere.* Doktori értekezés. Debreceni Egyetem

Őrsi, B., Kovács, J., & Csukonyi, C. (2024). Accepting a robot request contradicting a human instruction in the function of robot attitudes and level of interdependency. *Computers in Human Behavior Reports*, *14*, 100385.

Őrsi, B., Lipták, M & Csukonyi, Cs. (2021). A robotokkal kapcsolatos negatív attitűd- és szorongásmérő eszközök vizsgálata. *Alkalmazott Pszichológia, 21*(4), 77–100.

UNIVERSITY AND NATIONAL LIBRARY
UNIVERSITY OF DEBRECEN
H-4002 Egyetem tér 1, Debrecen
Phone: +3652/410-443, email: publikaciok@lib.unideb.hu

Registry number: DEENK/577/2024.PL
Subject: PhD Publication List

Candidate: Balázs Őrsi
Doctoral School: Doctoral School of Human Sciences
MTMT ID: 10070233

List of publications related to the dissertation

Hungarian scientific articles in Hungarian journals (2)

1. **Őrsi, B.**, Lipták, M. A., Csukonyi, C.: A robotokkal kapcsolatos negatív attitűd- és szorongásmérő eszközök vizsgálata.
   Alk. Pszichol. 21 (4), 77-100, 2021. ISSN: 1419-872X.
   DOI: http://dx.doi.org/10.17627/ALKPSZICH.2021.4.77
2. **Őrsi, B.**, Csukonyi, C.: A robotszorongás elméleti áttekintése.
   Psychiatr Hung. 35 (2), 175-181, 2020. ISSN: 0237-7896.

Foreign language scientific articles in Hungarian journals (1)

3. **Őrsi, B.**, Csukonyi, C.: Psychological Aspects and Opinions about Some Typical Robots and Robots in General.
   Recent Innov. Mechatron. 10 (1), 1-8, 2023. EISSN: 2064-9622.
   DOI: http://dx.doi.org/10.17667/riim.2023.03

Foreign language scientific articles in international journals (2)

4. **Őrsi, B.**, Kovács, J., Csukonyi, C.: Accepting a robot request contradicting a human instruction in the function of robot attitudes and level of interdependency.
   Computers in Human Behavior Reports. 14, 1-10, 2024. ISSN: 2451-9588.
   DOI: http://dx.doi.org/10.1016/j.chbr.2024.100385
   IF: 4.9 (2023)
5. Szabó, B., **Őrsi, B.**, Csukonyi, C.: Robots for surgeons? Surgeons for robots? Exploring the acceptance of robotic surgery in the light of attitudes and trust in robots.
   BMC Psychol. 12 (1), 1-11, 2024. EISSN: 2050-7283.
   DOI: http://dx.doi.org/10.1186/s40359-024-01529-8
   IF: 2.7 (2023)

UNIVERSITY of DEBRECEN

UNIVERSITY AND NATIONAL LIBRARY
UNIVERSITY OF DEBRECEN
H-4002 Egyetem tér 1, Debrecen
Phone: +3652/410-443, email: publikaciok@lib.unideb.hu

Foreign language conference proceedings (1)

6. **Őrsi, B.**, Csukonyi, C., Korondi, P.: Organic Human-Robot Interactions: Psychological Aspects to Help Social Robots Become Sociable.
In: 2024 IEEE/SICE International Symposium on System Integration (SII), IEEE, Piscataway, 1038-1044, 2024. ISBN: 9798350312072

List of other publications

Hungarian book chapters (2)

7. Bátfai, N., Csukonyi, C., **Őrsi, B.**, Papp, D.: Esport: a játékipar szent gráljai.
In: Az e-sport az élre tör : tematikus különszám. Szerk.: Bácsné Bába Éva , Balogh László, Szabados György Norbert, Ráthonyi Gergely Gábor, Harangi-Rákos Mónika, Lenténé Puskás Andrea, Biró Melinda, Debreceni Egyetem Sporttudományi Koordinációs Intézet Debreceni Egyetem, Sportgazdasági és - menedzsment Intézet, Debrecen, 70-80, 2021, (Válogatott tanulmányok a sporttudomány köréből, ISSN 2631-0910 ; 6)

8. Bátfai, N., Csukonyi, C., **Őrsi, B.**: Robotpszichológiai és filogenetikai videójátékfa.
In: Az e-sport az élre tör : tematikus különszám. Szerk.: Bácsné Bába Éva , Balogh László, Szabados György Norbert, Ráthonyi Gergely Gábor, Harangi-Rákos Mónika, Lenténé Puskás Andrea, Biró Melinda, Debreceni Egyetem Sporttudományi Koordinációs Intézet Debreceni Egyetem, Sportgazdasági és - menedzsment Intézet, Debrecen, 61-69, 2021, (Válogatott tanulmányok a sporttudomány köréből, ISSN 2631-0910 ; 6)

Hungarian scientific articles in Hungarian journals (5)

9. Új, E. D., Csukonyi, C., **Őrsi, B.**, Kiss, B.: A célkitűzés hatása az utánpótlás korú kosárlabdázók fejleszthetőségére a kontrollhely és motiváció forrásának tükrében.
Stadium Hung. J. Sport Sci. 4 (1), 1-20, 2021. ISSN: 2676-9506.
DOI: http://dx.doi.org/10.36439/shjs/2021/1/9504

10. Papp, D., **Őrsi, B.**, Csukonyi, C.: Digitális technológia a vezetésben.
Új Munkaügyi Szemle. 1 (1), 82-89, 2020. EISSN: 2677-1306.

11. Új, E. D., **Őrsi, B.**, Csukonyi, C.: Kockázatvállalási tendenciák a profi sportolóknál versus kockázatvállalás a munkahelyen.
Opus et educatio. 7 (3), 276-281, 2020. ISSN: 2064-9908.

12. **Őrsi, B.**: A mesterséges munkatársakról: gondolati előretekintés.
Munkaügyi szle. 5, 48-52, 2019. EISSN: 2064-3748.

13. Papp, D., **Őrsi, B.**, Csukonyi, C.: Digitális technológia a vezetésben.
Munkaügyi szle. 5 (12), 1-8, 2019. EISSN: 2064-3748.

UNIVERSITY of DEBRECEN

UNIVERSITY AND NATIONAL LIBRARY
UNIVERSITY OF DEBRECEN
H-4002 Egyetem tér 1, Debrecen
Phone: +3652/410-443, email: publikaciok@lib.unideb.hu

Foreign language scientific articles in international journals (1)

14. Kovács, K. E., Kovács, K., Szabó, F., Dan, B., Szakál, Z., Moravecz, M., Szabó, D., Olajos, T., Csukonyi, C., Papp, D., **Őrsi, B.**, Pusztai, G.: Sport Motivation from the Perspective of Health, Institutional Embeddedness and Academic Persistence among Higher Educational Students.
Int. J. Environ. Res. Public Health. 19 (12), 1-23, 2022. ISSN: 1661-7827.
DOI: http://dx.doi.org/10.3390/ijerph19127423

Foreign language conference proceedings (1)

15. Keczán, L., **Őrsi, B.**, Katona, K., Mikuska, R., Neamah, H. A., Csukonyi, C., Korondi, P.: Technical limitations of Organic Human-Robot Interaction (O-HRI) for mobile robots moving amongst humans.
In: 2024 IEEE 21st International Power Electronics and Motion Control Conference (PEMC), IEEE, [s.l.], 1-6, 2024. ISBN: 9798350385236

Hungarian abstracts (2)

16. Kovács, K. E., Szakál, Z., Bíró, Z., Kovács, M., **Őrsi, B.**: A sportból történő lemorzsoldódás elemzése serdülők és fiatalok körében - egy szisztematikus összefoglaló tanulmány tanulságai.
In: Az oktatás időszerű narratívumai : Absztraktkötet. Szerk.: Juhász Erika; Gyányi István, Magyar Nevelés- és Oktatáskutatók Egyesülete, Eger, 255-256, 2024. ISBN: 9786155657153

17. Szakál, Z., Bíró, Z., Kovács, M., **Őrsi, B.**, Kovács, K. E.: A sportperzisztencia támogatása - a sportból való lemorzsolódás prevenciós lehetőségei egy szisztematikus összefoglaló tanulmány tanulságai alapján.
In: XXIV. Országos Neveléstudományi Konferencia : Absztraktkötet : Oktatás és nevelés a társadalmi jóllét szolgálatában. A nevelés és az oktatás kihívásai a válságok korában. Szerk.: Bocsi Veronika, Csók Cintia, MTA Pedagógiai Tudományos Bizottság ; Debreceni Egyetem Gyermeknevelési és Gyógypedagógiai Kar ; Debreceni Egyetem Bölcsészettudományi Kar Nevelés- és Művelődéstudományok Intézete, Debrecen, 137-138, 2024. ISBN: 9789634906551

Total IF of journals (all publications): 7,6
Total IF of journals (publications related to the dissertation): 7,6

The Candidate's publication data submitted to the iDEa Tudóstér have been validated by DEENK on the basis of the Journal Citation Report (Impact Factor) database.

25 November, 2024