Justus-Liebig-University Giessen

Dissertation

From Aversion to Adoption: Exploring the Influence of Conversational Agents on Users

Submitted in fulfillment of the requirements for the degree of

DOCTOR RERUM POLITICARUM (Dr. rer. pol.)

at the

Justus Liebig University Giessen

Faculty of Economics and Business Studies

Chair for Digitalization, E-Business and Operations Management

by

Pascal Oliver Heßler

March 10, 2025

Druckdatum: 10.03.2025
URL: https://doi.org/10.22029/jlupub-19606

Justus-Liebig-Universität Gießen
Fachbereich Wirtschaftswissenschaften
Professur für Digitalisierung, E-Business und Operations Management
Licher Straße 74
35394 Gießen

Dekanin:	Prof. Dr. Corinna Ewelt-Knauer
Erstgutacherin:	Prof. Dr. Jella Pfeiffer
Zweitgutachter:	Prof. Dr. Alexander Haas

Acknowledgments

First, I want to express my gratitude to my supervisor, Prof. Dr. Jella Pfeiffer. This dissertation would not have been possible without her invaluable guidance and support. I am especially thankful for the freedom she granted me in my research, the opportunities she provided, the possibilities she enabled, her empathy, and her open-mindedness.

I also extend my sincere thanks to my second supervisor, Prof. Dr. Alexander Haas, who took over from Prof. Dr. Sebastian Hafenbrädl and supported me during the final stages of this journey. Speaking of Prof. Dr. Sebastian Hafenbrädl, I am deeply grateful to him for the opportunity to visit him in Barcelona and for his unwavering support, attentive ear, and engaging discussions.

Additionally, I am very grateful to the exceptional students, co-authors, and faculty members from various institutions who enriched my journey with their collaboration and companionship.

Lastly, I wish to express my gratitude to my family and friends for their unwavering support and belief in me throughout this endeavor.

"*If I have seen further, it is by standing on the shoulders of giants*."

– Isaac Newton (1675)

Contents

1 General Introduction

1.1 Relevance & Motivation

At the beginning of this dissertation, Conversational Agents (CAs) were on the rise, gaining attention as a promising technology with significant potential. Industries, especially in customer service, began adopting CAs to improve cost efficiency and enhance user experiences. However, early implementations still had notable limitations: they were highly task-specific, struggled with unanticipated or complex inputs, and were prone to errors in understanding and responding (Zubatiy et al., 2023). Despite these challenges, falling implementation costs and increasing accessibility, empowered businesses to develop their own chatbots, accelerating innovation and laying the foundation for broader adoption and exploration of more sophisticated applications (Yan et al., 2016).

The release of OpenAI's ChatGPT in November the 30th, 2022 (OpenAI, 2022), marked a pivotal moment for CAs. Within just two months, ChatGPT reached 100 million monthly active users, showcasing the immense public interest and usability of generative AI technologies (Reuters, 2023). ChatGPT–or Large Language Models (LLMs) in general–have been compared to other disruptive technologies such as the World Wide Web, cloud computing, and statistical machine learning (Fernandez et al., 2023), underlining their transformative potential. Unlike earlier systems, ChatGPT demonstrated the capacity to address a wide array of user queries. This shift from specialized tools to a more general-purpose conversational system highlighted the potential of LLMs to fundamentally change how humans interact with technology.

The growth of the generative AI sector underscores this trend. In 2020, the sector generated approximately $14 billion in revenue, representing less than 1% of the broader IT industry, including hardware, software, and gaming markets (Bloomberg, 2023). By 2024, revenue had surged to $137 billion, capturing 3% of the sector. According to Bloomberg, projections indicate that generative AI will generate $1.304 trillion in revenue by 2032, comprising 12% of the entire IT industry. While CAs represent only one facet of this expanding ecosystem, these numbers illustrate their growing importance.

Before the aforementioned generative AI advancements, systems like Alexa, Siri and Google Assistant had already familiarized end users with the concept of voice-based interaction and in general with CAs in general. Once limited in scope, CAs have become highly accessible and dramatically more powerful with the advent of LLMs, enabling rich, context-aware interactions. By making ChatGPT freely available, OpenAI ensured that virtually anyone with an Internet connection could access and experiment with the system. In parallel, major tech companies such as Microsoft (Copilot[1]), Apple (Apple Intelligence[2]), and

[1] https://copilot.microsoft.com/
[2] https://www.apple.com/apple-intelligence/

Alphabet (Gemini[3]) are integrating generative AI systems into their products, further embedding these technologies into everyday life.

As these technologies have grown more advanced and widely used, they have raised important questions about how people perceive and interact with them. Acceptance of algorithms, particularly in the form of conversational agents, is a critical issue. While the technical capabilities of CAs have evolved significantly, the human factors influencing their adoption remain a central concern. This dissertation addresses one of these key challenges: algorithm aversion.

Prior research has demonstrated that humans often distrust algorithms—a phenomenon known as "algorithm aversion" (Dietvorst et al., 2015). This aversion can stem from several causes, with one prominent reason being that algorithms are not human, and people generally prefer interactions with real humans (Dietvorst et al., 2015; Sinha & Swearingen, 2001). A clear advantage of human interaction is the relative ease of establishing trust, a quality often lacking in computer-based interactions, since it is harder to predict how they will act (Lankton et al., 2015). Trust is important when we are face with uncertainties (McKnight & Chervany, 2000). When judging if we can trust another human we can rely on multiple types of information, including non-verbal (e.g., facial expressions, body language) and verbal cues (e.g., voice) information. In contrast, when communicating with a CA this information exchange is limited, making it more difficult to judge the counterpart. Finally, people believe that computers are not able to make moral decisions, increasing the algorithm aversion in moral contexts (Bigman & Gray, 2018).

However, reality shows that humans frequently interact with computers as if they were human, applying human-to-human social scripts—a phenomenon known as anthropomorphism (Nass et al., 1994; Nass & Moon, 2000). Anthropomorphism involves attributing human traits to non-human entities (Epley et al., 2008). Even the first developed CA with a very simple interface was anthropomorphized (Weizenbaum, 1966).[4] This raises the intriguing possibility that anthropomorphization might serve as a potential solution to mitigate algorithm aversion.

Nevertheless, even when CAs are anthropomorphized, algorithm aversion may persist. Anthropomorphizing may alleviate some symptoms, but it does not address the root causes of the aversion. Therefore, it is critical to explore situations where algorithm aversion is more pronounced and where it may be less relevant. Research shows that different task types lead to varying levels of algorithm aversions (Castelo et al., 2019). But it is not only the task-type which influences this; often, the context itself defines the task type. In financial contexts decisions are mostly based on quantifiable and measurable facts (Castelo et al., 2019; Logg, 2017), while in more prosocial contexts those facts are also relevant but are paired with subjective criteria, complex ideas, personal preferences and emotions (Seeger

[3] https://gemini.google.com/
[4] The first CA was ELIZA developed by Joseph Weizenbaum in 1966.

et al., 2021; Waytz et al., 2014). This difference might create contexts where interacting with a computer might be preferable (Logg et al., 2019)—such as when discussing financial topics (e.g., in for-profit contexts) or considering socially sensitive or unethical topics. Conversely, in situations where empathy plays a central role (e.g., in prosocial contexts), the presence of a human counterpart may be essential.

1.2 Research question

This dissertation aims to address two distinct research questions. First, we ask:

1. Is the acceptance of algorithms different in prosocial vs for-profit context, and why might this distinction exist?

This question explores whether context influences algorithm acceptance. This research contrasts two opposing scenarios: one where the primary goal is making money (for-profit) (Haas et al., 2014) and one where the focus is helping others (prosocial), promoting prosocial or rather altruistic motives (Galak et al., 2011; Haas et al., 2014).

Our findings revealed a consistent preference for human interaction in both scenarios, though the preference was significantly stronger in the prosocial context, leading to our second research question:

2. How do different anthropomorphized CAs influence user behavior in prosocial contexts?

Since previous research has already established that anthropomorphization increases trust (Lankton et al., 2015) and reduces algorithm aversion (Waytz et al., 2014), our focus shifted towards understanding its downstream effects. Specifically, we explored whether anthropomorphization influences user behavior and especially how it affects user in prosocial contexts. Do people behave differently when interacting with a more human-like CA compared to a computer-like one? For instance, do they donate more to charities after engaging with a human-like CA, or is their generosity unaffected by the agent's design? How do prosocial contexts shift the expectations about CAs and how does this affect user behavior? These are some of the questions this dissertation seeks to address.

1.3 Method

This dissertation is firmly rooted in the interdisciplinary field of Information Systems, combining insights and theories from psychology, computer science, and related disciplines to create a robust theoretical foundation.

To address our research questions, we primarily employed quantitative experimental methods. Across all studies, we conducted experiments and gathered data via crowdworking platforms such as MTurk and Prolific. Survey development played a crucial role in our research process. In the first study, interactions were simulated using only static screenshots. In the two subsequent studies, we developed purpose-built

CAs using Microsoft Azure's botframework[5], allowing participants to interact directly with the agents through a custom-built website designed with the Python framework Django and later the JavaScript framework React. For the final study, we utilized the OpenAI API to integrate a more advanced CA, enabling richer and more dynamic interactions. From a data analysis perspective, the complexity of our research models dictated our methodological approach. We utilized a range of techniques, starting with basic regressions for simpler models and advancing to more sophisticated methods for intricate analyses. These included seemingly unrelated regressions (SUREG) as well as structural equation models (SEM). For SEM, we applied both partial least squares (PLS-SEM) and covariance-based (CB-SEM) approaches, depending on the specific requirements of each study. Additionally, power analyses were conducted to ensure that our studies were adequately powered to detect meaningful effects.

1.4 Structure & Contribution

Building on decades of research, this dissertation addresses critical questions about human interactions with CAs. By examining how people perceive and are influenced by these systems, it aims to deepen our understanding of these interactions and their broader societal implications. This research seeks to provide valuable insights into the impact of this rapidly evolving technology.

In the following I will briefly summarize each of the four papers of this dissertation, highlighting their research questions and central contributions. Finally, I will summarize the contributions of my dissertation.

The first paper addresses the initial research question of the dissertation: How and why do people exhibit different levels of algorithm aversion across various contexts? Specifically, it examines when individuals prefer human-like decision support systems over computer-like ones and why these preferences differ across contexts. Grounded in the self-humanization framework (Haslam et al., 2005; Haslam, 2006), the study hypothesizes that prosocial contexts, such as microlending without financial returns, elevate the importance of autonomy and empathy, both of which conflict with algorithmic decision support systems. This tension is theorized to drive higher algorithm aversion in prosocial settings compared to for-profit contexts, where financial incentives are more dominant (Haas et al., 2014). The research further investigates whether anthropomorphization of decision support systems can mitigate this aversion by addressing the conflict between users' need for autonomy and empathy and the nature of algorithmic assistance.

This paper contributes to both theory and practice by introducing self-humanization as a novel lens for understanding contextual differences in algorithm aversion. Grounded in this framework, the study highlights autonomy and empathy as crucial mechanisms connecting context to user behavior, and thus extending traditional models that emphasize ease of use and perceived usefulness. Emphasizing self-

[5] https://github.com/Microsoft/botbuilder-python

humanization explains how prosocial contexts, such as microlending without financial returns, elevate the importance of autonomy and empathy, both of which conflict with algorithmic decision support systems. The study underscores the importance of considering these factors in the design of digital platforms, particularly in prosocial contexts. On the practical side, the findings provide guidance for designing decision support systems that reduce algorithm aversion through human-centric features like anthropomorphization and emotion detection. Such decision support systems have the potential to not only improve user satisfaction but also foster prosocial behaviors, thereby enhancing both individual and societal outcomes.

The second paper addresses the second research question: How do different anthropomorphized CAs influence user behavior in prosocial contexts? To broaden our scope, we include a control group, and refer to both anthropomorphized CAs and the control as Assistant Systems (AS). This broader scope allows us to compare the influence of a CA against systems without one. While existing research often focuses on the impact of individual social cues, such as names or emojis (e.g., Feine et al., 2019), this study extends the discussion by incorporating multiple cues, including voice, which conveys human-like attributes such as emotion and identity (Scherer, 1995; Stern et al., 2021). In doing so, it investigates both the positive and potential negative effects of anthropomorphism through social presence. While social presence is typically associated with positive outcomes, such as increased trust and empathy (Epley et al., 2007; Qiu & Benbasat, 2009), this study hypothesizes that it may also evoke the *feeling of being observed*, either by others or by the CA itself. This heightened sense of observation, as shown in prior research (Rhim et al., 2022), is expected to prompt more socially acceptable behavior, leading to increased investments in prosocial projects when social presence is higher.

This study contributes to the understanding of anthropomorphized ASs by investigating how different social cues, particularly voice, influence user behavior in microlending. Contrary to our expectations, voice did not significantly increase perceived anthropomorphism compared to a text-based AS, highlighting the complexity of this relationship. This *feeling of being observed*, surprisingly, negatively impacts investment levels, offering new insights into the potential drawbacks of increased social presence in decision-making environments. Our findings further suggest that designing microlending platforms to encourage users to explore multiple projects could increase prosocial investments. Overall, this research emphasizes the need for context-sensitive design and nuanced approaches to using social cues in ASs.

The third paper directly builds on the second one. As mentioned, we collected data in an online experiment utilizing a human voice to increase anthropomorphization. In an exploratory analysis, we found that male participants reacted much more strongly to the female voice of the CA. This paper expands on these findings by examining how user gender interacts with CA gender to influence preferences in pro-social and for-profit decisions. Grounded in research on gender stereotypes and human-computer interactions (Heilman, 2012; Nass & Moon, 2000), it explores whether these stereotypes shape user behavior when interacting with gendered CAs. By incorporating both male and female, with

and without voice CAs in a new experiment the paper aims to provide a nuanced understanding of the interaction between genders

The findings of this study reveal that while a CA's gender does not directly influence user decisions, the presence of a human voice significantly impacts behavior, particularly among men. This suggests that the voice itself, rather than gender stereotypes, plays a pivotal role in shaping user interactions. However, the study underscores the harm caused by relying on gender stereotypes in CA design, as these practices risk reinforcing societal biases without necessarily enhancing functionality. By demonstrating the limited practical benefits of gendered personas and highlighting the ethical challenges of such designs, this research advocates for a shift toward inclusive and stereotype-free CA development. These insights contribute to a growing body of literature that calls for rethinking gendered CAs, particularly in light of evolving technologies like LLMs, which often amplify existing biases.

The fourth and final paper builds on the recent developments in generative AI, particularly focusing on the impact of anthropomorphizing CAs like ChatGPT on prosocial decision-making. This study leverages the Stereotype Content Model (SCM) from psychology (Cuddy et al., 2008; Fiske et al., 2002, 2007; Khadpe et al., 2020), which posits that individuals are judged on competence and warmth, with warmth being prioritized. By integrating the self-humanization framework, we hypothesize that computer-like CAs are perceived as competent but cold, while human-like CAs are seen as warm. However, the definition of competence in this context remains ambiguous, as being human-like could also enhance perceptions of competence, particularly in prosocial scenarios. We examine how these two dimensions mediate the relationship between anthropomorphism and algorithm aversion, expecting warmth to play a central role in influencing donation behavior.

Our primary contribution is the empirical finding that competence—rather than warmth—plays the dominant role in reducing algorithm aversion and increasing donation amounts in the context of charities. While we initially hypothesized that both competence and warmth would mediate the effects of anthropomorphism, our results show that warmth does not significantly impact donation behavior or algorithm aversion. This insight challenges the assumption that warmth is the key factor in anthropomorphizing CAs and suggests that future research should focus on the importance of competence in designing CAs for prosocial contexts. Furthermore, our findings indicate that both human-like and computer-like CAs elicit similar levels of algorithm aversion, highlighting the complexity of anthropomorphization and its effects. Finally, we demonstrate that it is possible to manipulate ChatGPT to exhibit different behaviors, a finding with significant implications for research, as it enables the testing of various manipulations.

This dissertation explores how anthropomorphized CAs influence user behavior in prosocial decision-making contexts. The first paper introduces the self-humanization framework to understand algorithm aversion across different contexts, emphasizing the role of autonomy and empathy. The second paper highlights the complex effects of social cues, especially voice, on user behavior, introducing the concept

of feeling observed. The third paper investigates the potential matching-gender effect of voice actors and participants, highlighting that voice, not gender, is the driving force. The fourth paper deepens our understanding of how competence, rather than warmth, plays a central role in influencing algorithm aversion and prosocial decisions. Together, these studies advance our understanding of human-CA interactions and provide insights into designing effective, context-sensitive decision support systems.

The value of this dissertation lies not only in the individual contributions of each paper but also in the multiplicity of perspectives they offer. Exploring different facets of anthropomorphized CAs–sometimes yielding complementary and, at other times, contradictory findings–underscores the importance of diverse approaches in research. This multiplicity allows us to see further than any single perspective could, offering a richer and more nuanced understanding of complex phenomena and paving the way for more holistic solutions in the design of effective, context-sensitive CAs.

1.5 References

Bigman, Y. E., & Gray, K. (2018). People are averse to machines making moral decisions. Cognition, 181, 21–34. https://doi.org/10.1016/j.cognition.2018.08.003

Bloomberg. (2023, June 1). Generative AI to Become a $1.3 Trillion Market by 2032. Bloomberg L.P. https://www.bloomberg.com/company/press/generative-ai-to-become-a-1-3-trillion-market-by-2032-research-finds/

Castelo, N., Bos, M. W., & Lehmann, D. R. (2019). Task-dependent algorithm aversion. Journal of Marketing Research, 56(5), 809–825. https://doi.org/10.1177/0022243719851788

Cuddy, A. J. C., Fiske, S. T., & Glick, P. (2008). Warmth and Competence as Universal Dimensions of Social Perception: The Stereotype Content Model and the BIAS Map. In Advances in Experimental Social Psychology (Vol. 40, pp. 61–149). Elsevier. https://doi.org/10.1016/S0065-2601(07)00002-0

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. Journal of Experimental Psychology: General, 144(1), 114–126. https://doi.org/10.1037/xge0000033

Epley, N., Waytz, A., Akalis, S., & Cacioppo, J. T. (2008). When we need a human: Motivational determinants of anthropomorphism. Social Cognition, 26(2), 143–155. https://doi.org/10.1521/soco.2008.26.2.143

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. Psychological Review, 114(4), 864–886. https://doi.org/10.1037/0033-295X.114.4.864

Feine, J., Gnewuch, U., Morana, S., & Maedche, A. (2019). A taxonomy of social cues for conversational agents. International Journal of Human-Computer Studies, 132, 138–161. https://doi.org/10.1016/j.ijhcs.2019.07.009

Fernandez, R. C., Elmore, A. J., Franklin, M. J., Krishnan, S., & Tan, C. (2023). How Large Language Models Will Disrupt Data Management. Proceedings of the VLDB Endowment, 16(11), 3302–3309. https://doi.org/10.14778/3611479.3611527

Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. Trends in Cognitive Sciences, 11(2), 77–83. https://doi.org/10.1016/j.tics.2006.11.005

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. Journal of Personality and Social Psychology, 82(6), 878–902. https://doi.org/10.1037/0022-3514.82.6.878

Galak, J., Small, D., & Stephen, A. T. (2011). Microfinance decision making: A field study of prosocial lending. Journal of Marketing Research, 48(SPL), 130137. https://doi.org/10.1509/jmkr.48.SPL.S130

Haas, P., Blohm, I., & Leimeister, J. M. (2014). An empirical taxonomy of crowdfunding intermediaries. In Proceedings of the International Conference on Information Systems—Building a Better World through Information Systems. Association for Information Sytems.

Haslam, N. (2006). Dehumanization: An integrative review. Personality and Social Psychology Review, 10(3), 252–264. https://doi.org/10.1207/s15327957pspr1003_4

Haslam, N., Bain, P., Douge, L., Lee, M., & Bastian, B. (2005). More human than you: Attributing humanness to self and others. Journal of Personality and Social Psychology, 89(6), 937–950. https://doi.org/10.1037/0022-3514.89.6.937

Heilman, M. E. (2012). Gender stereotypes and workplace bias. *Research in Organizational Behavior*, *32*, 113–135. https://doi.org/10.1016/j.riob.2012.11.003

Khadpe, P., Krishna, R., Fei-Fei, L., Hancock, J. T., & Bernstein, M. S. (2020). Conceptual metaphors impact perceptions of human-AI collaboration. Proceedings of the ACM on Human-Computer Interaction, 4(CSCW2), 1–26. https://doi.org/10.1145/3415234

Lankton, N., McKnight, D. H., & Tripp, J. (2015). Technology, humanness, and trust: Rethinking trust in technology. Journal of The Association for Information Systems, 16(10), 880–918. https://doi.org/10.17705/1jais.00411

Logg, J. M. (2017). Theory of Machine: When Do People Rely on Algorithms? Harvard Business School Working Paper Series # 17-086. https://dash.harvard.edu/handle/1/31677474

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. Organizational Behavior and Human Decision Processes, 151, 90–103. https://doi.org/10.1016/j.obhdp.2018.12.005

McKnight, D. H., & Chervany, N. L. (2000). What is trust? A conceptual analysis and an interdisciplinary model.

Moon, J.-W., & Kim, Y.-G. (2001). Extending the TAM for a World-Wide-Web context. *Information & Management*, *38*(4), 217–230. https://doi.org/10.1016/S0378-7206(00)00061-6

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. Journal of Social Issues, 56(1), 81103. https://doi.org/10.1111/0022-4537.00153

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. In B. Adelson (Ed.), Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM. https://doi.org/10.1145/191666.191703

OpenAI. (2022, November 30). Introducing ChatGPT. https://openai.com/index/chatgpt/

Qiu, L., & Benbasat, I. (2009). Evaluating Anthropomorphic Product Recommendation Agents: A Social Relationship Perspective to Designing Information Systems. Journal of Management Information Systems, 25(4), 145–182. https://doi.org/10.2753/MIS0742-1222250405

Reuters. (2023, February 2). ChatGPT sets record for fastest-growing user base—Analyst note | Reuters. https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/

Rhim, J., Kwak, M., Gong, Y., & Gweon, G. (2022). Application of humanization to survey chatbots: Change in chatbot perception, interaction experience, and survey data quality. Computers in Human Behavior, 126, 107034. https://doi.org/10.1016/j.chb.2021.107034

Scherer, K. R. (1995). Expression of emotion in voice and music. Journal of Voice, 9(3), 235–248. https://doi.org/10.1016/S0892-1997(05)80231-0

Seeger, A.-M., Pfeiffer, J., & Heinzl, A. (2021). Texting with humanlike conversational agents: Designing for anthropomorphism. Journal of The Association for Information Systems, 22(4), 931–967. https://doi.org/10.17705/1jais.00685

Sinha, R., & Swearingen, K. (2001). Comparing recommendations made by online systems and friends. In Alan F. Smeaton & Jamie Callan (Eds.), DELOS. ERCIM.

Stern, J., Schild, C., Jones, B. C., DeBruine, L. M., Hahn, A., Puts, D. A., Zettler, I., Kordsmeyer, T. L., Feinberg, D., Zamfir, D., Penke, L., & Arslan, R. C. (2021). Do voices carry valid information about a speaker's personality? Journal of Research in Personality, 92, 104092. https://doi.org/10.1016/j.jrp.2021.104092

Waytz, A., Heafner, J., & Epley, N. (2014). The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. Journal of Experimental Social Psychology, 52, 113–117. https://doi.org/10.1016/j.jesp.2014.01.005

Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. Communications of the ACM, 9(1), 36–45. https://doi.org/10.1145/365153.365168

Yan, M., Castro, P., Cheng, P., & Ishakian, V. (2016). Building a Chatbot with Serverless Computing. Proceedings of the 1st International Workshop on Mashups of Things and APIs, 1–4. https://doi.org/10.1145/3007203.3007217

Zubatiy, T., Mathur, N., Heck, L., Vickers, K. L., Rozga, A., & Mynatt, E. D. (2023). “I don’t know how to help with that”—Learning from Limitations of Modern Conversational Agent Systems in Caregiving Networks. Proceedings of the ACM on Human-Computer Interaction, 7(CSCW2), 1–28. https://doi.org/10.1145/3610170

Table 1.1: Overview of Dissertation Papers

Paper	**When Self-Humanization Leads to Algorithm Aversion: What Users Want from Decision Support Systems on Prosocial Microlending Platforms**
Co-authors	Jella Pfeiffer and Sebastian Hafenbrädl
Status	Published in *Business & Information Systems Engineering*
Key contributions	My contribution is 80%: • Conceptualization of the theory and designing of the experiment (together with co-authors) • Collection of data via MTurk • Preparation and analysis of the data • Writing the manuscript (together with co-authors)
Paper	**Conversational Agents with Voice: How Social Presence Influences the User Behavior in Microlending Decisions**
Co-authors	Jella Pfeiffer and Matthias Unfried
Status	Published in Proceedings *European Conference on Information Systems* (2023)
Key contributions	My contribution is 75%: • Conceptualization of the theory and designing of experiment (together with co-authors) • Developing the experiment • Collection of data via Prolific • Preparation and analysis of the data • Writing the manuscript (together with co-authors)
Paper	**The Voice Effect: Rethinking Gender Stereotypes in Conversational Agent Design**
Co-authors	Jella Pfeiffer and Matthias Unfried
Status	Finished; not previously published.
Key contributions	My contribution is 90%: • Conceptualization of the theory • Developing the experiment (together with co-authors) • Collection of data via Prolific (together with co-authors) • Preparation and analysis of the data • Writing the manuscript
Paper	**Competence Over Warmth in Charitable Giving: The Algorithm Aversion Paradox of Humanizing Conversational Agents**
Co-authors	Jella Pfeiffer and Sebastian Hafenbrädl
Status	Working Paper
Key contributions	My contribution is 80%: • Conceptualization of the theory and designing of experiment (together with co-authors) • Developing the experiment (together with co-authors) • Collection of data via Prolific • Preparation and analysis of the data (together with co-authors) • Writing the manuscript

2 Paper A: When Self-Humanization Leads to Algorithm Aversion

What Users Want from Decision Support Systems on Prosocial Microlending Platforms

Pascal Oliver Heßler • Jella Pfeiffer • Sebastian Hafenbrädl

Abstract

Decision support systems are increasingly being adopted by various digital platforms. However, prior research has shown that certain contexts can induce algorithm aversion, leading people to reject their decision support. This paper investigates how and why the context in which users are making decisions (for-profit versus prosocial microlending decisions) affects their degree of algorithm aversion and ultimately their preference for more human-like (versus computer-like) decision support systems. The study proposes that contexts vary in their affordances for self-humanization. Specifically, people perceive prosocial decisions as more relevant to self-humanization than for-profit contexts, and, in consequence, they ascribe more importance to empathy and autonomy while making decisions in prosocial contexts. This increased importance of empathy and autonomy leads to a higher degree of algorithm aversion. At the same time, it also leads to a stronger preference for human-like decision support, which could therefore serve as a remedy for an algorithm aversion induced by the need for self-humanization. The results from an online experiment support the theorizing. The paper discusses both theoretical and design implications, especially for the potential of anthropomorphized conversational agents on platforms for prosocial decision-making.

2.1 Introduction

Decision support systems are becoming faster, smarter, and more powerful by the minute, and thus it is for good reason that they can be found on just about any successful internet platform in the form of recommendation systems, conversational agents, or interactive decision aids (Aggarwal 2016; Jung et al. 2018; Maedche et al. 2019; Pfeiffer et al. 2014). As these decision support systems spread to more and more domains of life, however, the question arises as to what extent users are willing to use them in every context. Indeed, while algorithms are being rapidly adopted in some contexts, prior research has shown that people are often algorithm averse (Castelo et al. 2019; Dietvorst et al. 2015) and that they might prefer the support of another human (Dietvorst et al. 2015; Sinha and Swearingen 2001; Yeomans et al. 2019)—for instance, if they perceive a task to be more subjective and thus requiring intuition as well as personal interpretation (Castelo et al. 2019; Inbar et al. 2010). In a related stream of research, Seeger et al. (2021) proposed that some tasks are more human-like, meaning that the support system is substituting for a human interaction partner and that this might affect users' expectations of the system's design. Overall, given the huge potential of decision support systems to facilitate and improve decision-making, there is continued interest in the question about context-specific reasons for algorithm aversion, both theoretically, as such reasons are closely tied to a deep understanding of its

underlying mechanisms, and practically, with an eye toward building context-specific remedies to overcome this bias.

In this paper, we address this question by building on the theoretical framework of self-humanization as a particularly suitable conceptual lens for explaining contextual differences in algorithm aversion. The central tenet of this framework is that people want to be seen by others, and to see themselves, as fully human (Haslam et al. 2005). In order to feel human, people place great importance on using abilities that have been called human nature attributes. They believe these attributes cannot be shared with machines—think, for example, of emotional responsiveness, interpersonal warmth, agency, cognitive openness, and depth (Haslam 2006). The main thesis of this paper is thus that in decision contexts where people see such human nature attributes as particularly important, they become algorithm averse and would prefer to be supported by a human (as humans have these attributes, but algorithms cannot possess them; Haslam (2006)). Specifically, we propose that the underlying reasons that make people averse to algorithms in contexts they deem relevant for self-humanization are two facets of self-humanization: the importance of empathy and autonomy.

In addition to this theoretical contribution, we consider the practical implications of our research model and propose decision support systems that imitate human-like characteristics as a remedy for humanization-induced algorithm aversion. We call such decision support systems human-like decision support. Imagine anthropomorphized conversational agents who, using natural language, emulate human-to-human communication (Maedche et al. 2019; Schuetzler et al. 2014; Seeger et al. 2021), or consider the applications of neurophysiological measurements for making communication between humans and computers emotionally richer (Picard 2003; Zheng and Lu 2015). Or contemplate the attempts made to compel black box artificial intelligence algorithms to explain their decisions to the user (Adadi and Berrada 2018; Barredo Arrieta et al. 2020). Such decision support systems are not only rising in popularity; they also prompt the user to ascribe human nature attributes to them.

One decision context for which human nature attributes are seen as particularly important is that of prosocial decisions, defined as decisions to benefit others. People decide to help others in need, volunteer for good causes, and give money to charities. Prior research has shown that two factors stemming from these human nature attributes are particularly relevant for prosocial decisions: empathy and autonomy. Indeed, rather than trying to rationally find the option that produces the maximal benefit to others (or, more generally, the maximal welfare gain) or delegating their decision to an algorithm that could approximate such rationality, people often prefer to actively and autonomously choose options aligned with their own subjective preferences (Berman et al. 2018). They aim to select options that feel right (i.e., that give them a warm glow (Andreoni 1990; Dunn et al. 2014)) and that allow them to experience empathy with the beneficiary (Galak et al. 2011; Loewenstein and Small 2007). Despite these rather peculiar characteristics of the prosocial decision context, people increasingly use digital platforms to engage in prosocial behavior (e.g., Galak et al. 2011). To the best of our knowledge, no previous research

has used the lens of self-humanization to illuminate the context of prosocial decisions with the aim of exploring how it explains algorithm aversion and developing domain-specific remedies.

From a research design perspective, prosocial decisions are also a particularly well-suited context for studying self-humanization and algorithm aversion. One type of prosocial platform, prosocial microlending, has a for-profit counterpart in the form of regular for-profit microlending platforms. On both types of platforms, users select entrepreneurs, with the main difference that on one platform, the users receive no interest payments and follow prosocial motives (Galak et al. 2011; Haas et al. 2014), whereas on the other, they want to make money (Haas et al. 2014). Comparing for-profit with prosocial microlending decisions allows us to change the context (and thereby the relevance of self-humanization) while keeping most elements of the decision process constant and thus to isolate the effect of the decision context as thoroughly as possible. Specifically, in an online experiment, we manipulated the relevance of self-humanization by randomly assigning participants to make decisions either on a for-profit or on a prosocial microlending platform.

Our experiment provides evidence supporting our research model and thus the hypothesized causal relationship between factors that have rarely been studied together and are important for understanding contextual differences in algorithm aversion. We thereby make three main contributions: First, we build on the theoretical lens of self-humanization to understand how differences between decision contexts (and with them different types of platforms) affect algorithm aversion. Second, we propose and test the two main mechanisms of how self-humanization drives this context-specific algorithm aversion: the importance the user gives to autonomy and the importance the user gives to empathy. Third, we explore the practical implications for how this context-specific algorithm aversion can be remedied: by making the decision support system appear more human-like. This solution obviously has direct implications for the designers of decision support systems in different contexts. Creating a decision support system based on these ideas carries the promise of not only satisfying users' desire to feel more human but also of reinforcing prosocial behavior. Overall, our results strongly support the idea that decision support systems cannot merely be copied and pasted between contexts but need to be thoroughly adapted to users' preferences and expectations to prevent and overcome algorithm aversion.

2.2 Theory

2.2.1 Algorithm Aversion

Algorithms have long been proposed as a means to overcome the cognitive limitations of humans (Burton et al. 2020; Dawes 1979; Meehl 1954). Indeed, several studies in different contexts have shown that algorithms can and do outperform humans, for example, in forecasting tasks (Grove et al. 2000) and supply chain distribution (Validi et al. 2015). While some form of algorithm appreciation seems to exist in some domains (Logg et al. 2019; Prahl and van Swol 2017), in many contexts, people seem to be

intuitively averse to using them, a phenomenon that was termed "algorithm aversion" by Dietvorst et al. (2015).

One initial focus of this research was users' high expectations concerning the performance of algorithms: they expect them to be perfect. Consequently, people quickly lose trust in algorithms once they see them err (Dietvorst et al. 2015, 2018). Of course, predicting the future perfectly is inherently difficult; thus, even extremely well-crafted algorithms will err from time to time (Dietvorst et al. 2015; Prahl and van Swol 2017). However, there are also cases in which people did not observe the algorithm, thus they could not learn about algorithmic failures (e.g., Longoni et al. 2019), and yet they still felt algorithm aversion. Taking the breadth of the phenomenon into account, we follow Jussupow et al. (2020) and define algorithm aversion as the "biased assessment of an algorithm which manifests in negative behaviors and attitudes towards the algorithm compared to a human agent" (p. 4).

Algorithm aversion is an umbrella term, and there are several different reasons underlying this biased assessment. In recent literature reviews, these different causes have been discussed and categorized (see, for example, Burton et al. (2020) and Jussupow et al. (2020)). Let us give a few examples: We already mentioned the expectation that an algorithm should work perfectly (Dietvorst et al. 2015, 2018), which fits into the larger category of users' beliefs about what an algorithm is capable of, and which might be driven by a user's domain-specific expertise—with experts often showing higher degrees of algorithm aversion. Relatedly, humans make decisions differently from the way computers do (cognitive compatibility), for instance, by using heuristics, which are simple decision strategies that ignore part of the available information (Hafenbrädl et al. 2016; Hoffrage et al. 2018). Most of the time humans act in a world of uncertainty where not all possible consequences (and their probabilities) of a decision are known or knowable (Neth and Gigerenzer 2015), whereas typically algorithms optimize under risk, which means they implicitly assume that they have all outcomes and probabilities (divergent rationalities). Moreover, the category of decision autonomy describes the feeling of being in control, which could be diminished if one cannot understand how an algorithm actually makes decisions, or if one cannot influence and control how an algorithm makes the decision. As these examples illustrate, there are many different categories of causes for algorithm aversion; it comes in many flavors, forms, and functions. Some of these causes are driven not only by features of the algorithms themselves, but also by features of the context in which the algorithms are used (Castelo et al. 2019).

2.3 Self-Humanization

One theoretical dimension that seems particularly relevant for explaining contextual differences in algorithm aversion stems from the theoretical framework of self-humanization (Haslam 2006). Haslam et al. (2005) proposed that people want to be seen by others, and to see themselves, as fully human—to the point where they see themselves as more human than others. There are two distinct senses of humanness that contribute to being seen as fully human (Haslam 2006). First, uniquely human attributes

distinguish humans from animals, although despite being labeled uniquely human (e.g., cognitive capabilities, like logic, and rationality), they can be shared with machines. Second, and more importantly for explaining algorithm aversion, human nature attributes comprise attributes that people (across cultures) believe cannot be shared with machines (although potentially with animals), and thus, by extension, with algorithms and decision support systems (Haslam et al. 2008; Kahn et al. 2006). In his review paper, Haslam (2006) proposed five categories of human nature attributes: emotional responsiveness, interpersonal warmth, agency, cognitive openness, and depth. Prior research has found that people assess themselves (relative to others) to more strongly embody human nature attributes, especially openness, warmth, and emotionality (Haslam et al. 2005). Moreover, they also want to see themselves, and be seen by others, as possessing human nature attributes, which are perceived as more important and more deeply rooted in the person, relative to uniquely human attributes (Bain et al. 2006; Haslam et al. 2000; Haslam et al. 2004).

2.3.1 Overcoming Algorithm Aversion with Human-like Decision Support

In sum, the theoretical lens of self-humanization highlights that in contexts that people deem relevant for their self-humanization, there is a fundamental tension between human nature attributes on the one hand and using machines, algorithms, and computerized decision support systems on the other. Yet, when it comes to decisions on digital platforms, the sheer number of possibilities on many platforms can be overwhelming, and users long for ways to reduce the decision effort (e.g., Häubl and Trifts 2000). In principle, this renders the superior capabilities of algorithms to screen and integrate large amounts of information very attractive. The question arises whether it is possible, and if so, how, to make algorithm-based decision support more palatable to decision-makers in contexts in which they experience self-humanization-driven algorithm aversion.

The theoretical lens of self-humanization not only allows for understanding the underlying reasons for algorithm aversion in such contexts but also points to a potential solution: create the impression that the decision support is more human-like (and less machine-like). The intuitive classification of ways to support decisions in human-like and computer-like decision support stems from Seeger et al. (2021) and their concept of human-like versus computer-like tasks in the context of conversational agents. Human-like decision tasks are tasks in which a conversational agent is substituted for a human interaction partner (Lankton et al. 2015). These are tasks that are typical for a human (Seeger et al. 2021). We adapted the definition of human-like versus computer-like tasks from Seeger et al. (2021) and tailored it to decision support systems: human-like decision support refers to a decision support system that has characteristics that are typical for a human (e.g., possessing human nature attributes). As there is not necessarily a clear separation between these types of decision support systems, they can be placed on a continuum (Lankton et al. 2015).

There are multiple ways to dress up a decision support system to make it come across as more human-like. One prominent approach relies on anthropomorphization, which literally means humanizing (Epley et al. 2007), for example, through the use of social cues (Gnewuch et al. 2017; Seeger et al. 2021). A further way to create more human-like decision support systems might be to let computers simulate emotions, which is a burgeoning research area in computer science (e.g., affective computing). Decision support systems could try to give the user the impression that the computer has feelings by letting the computer detect emotions in both users and loan recipients with algorithms (Swangnetr and Kaber 2013). For example, the decision support system could try to infer emotions from the recipient's picture (Garcia-Garcia et al. 2017) or from text using sentiment analysis (Yadollahi et al. 2017), and a virtual agent might even be able to assume different facial expressions (Gordon et al. 2019).[6]

2.3.2 For-Profit versus Prosocial Decision Contexts

One straightforward operationalization of such contextual differences in the relevance of human nature attributes is to compare and contrast for-profit with prosocial decision-making contexts. For-profit decisions are defined as decisions people make to make money (e.g., interest payments from entrepreneurs), whereas prosocial decisions are defined as decisions that people make for the benefit for others (Eisenberg and Miller 1987).

We chose the domain of microlending decisions because there are for-profit and prosocial versions of microlending platforms, which creates a natural comparison that allows us to experimentally manipulate the context of a digital platform as cleanly as possible. Microlending itself is a relatively new financial instrument for providing entrepreneurs with small loans when traditional sources of financing may be unobtainable for them, for instance, due to their lack of collateral (Allison et al. 2013; Bruton et al. 2011). Peer-to-peer online platforms feature an emerging form of microlending that allows individuals to select entrepreneurs on the basis of the information contained in investment profiles, for instance, on for-profit microlending platforms like Prosper, FundedByMe, and Wisefund and on prosocial platforms like KIVA, GoFundMe, and Lend for Peace.

Because of for-profit decisions' strong focus on making money (Haas et al. 2014), the challenge of decision-making in such a context amounts to making good inferences about which loans will likely be paid back on time or even be paid back at all (Moss et al. 2015). In prosocial microlending, in contrast, lenders want to help someone in need (e.g., small business owners in developing countries) by lending them money interest-free (e.g., Allison et al. 2015; Galak et al. 2011). Prior research has found that in prosocial contexts, people do not think (or at least they act as if they do not think) that options can be objectively ranked (Berman et al. 2018), and thus they believe that there is no objectively best option that would likely have the largest positive impact on social welfare overall (Caviola et al. 2020).

[6] For practical examples, see the AI Companion from Luka https://replika.ai/ or Kuki AI from Pandora https://www.kuki.ai/.

Consequently, people prefer to base their decisions on more subjective factors—which are typically related to and driven by the abilities of human nature described above (e.g., their experience of empathy). Relying on their human nature capability to connect with the beneficiary in prosocial microlending thus provides an ideal contrast to the clearly defined criteria that lend themselves to rational optimization by machines in for-profit microlending (Bruton et al. 2011; Moss et al. 2015).

2.3.3 Human Nature Attributes, Empathy, and Autonomy

What are the implications of the particular relevance of human nature attributes for explaining contextual differences in algorithm aversion? The first factor stemming from these human nature attributes that is particularly relevant for prosocial decision-making is empathy, which is defined as the ability to take the emotional perspective of someone else—feeling as others—and includes the feeling of sympathy—feeling for others (Batson 2014; Cuff et al. 2016; Davis 1983; Loewenstein and Small 2007). Emotions in general, and empathy in particular, play a crucial role when making prosocial decisions (Barasch et al. 2014; Berman et al. 2018; Caviola et al. 2020). For instance, in the process of scrutinizing potential recipients of prosocial lending—that is, browsing through a list of entrepreneurs in need—people will emotionally react to photos and individual stories and often ultimately make their decisions based on this empathic reaction (Barasch et al. 2014; Eisenberg and Miller 1987; Herzenstein et al. 2011). Prompting people to adopt a more deliberative information processing approach (thus reducing their reliance on empathy) has been found to lower donations for recipients (Dickert et al. 2011; Small et al. 2007). Galak et al. (2011) provided evidence that in prosocial decisions, because similarity reduces social distance and facilitates empathy, people spend more money to help those who are more similar to themselves. More broadly, feelings of empathy and sympathy (as well as emotions such as fear, guilt, pity—cf. Sargeant et al. (2006)—and regret—Martinez et al. (2011)) feature prominently among the factors that influence how much people are willing to give (Galak et al. 2011; Hamilton and Sherman 1996; Pavey et al. 2012).

The second factor stemming from these human nature attributes that is particularly relevant for prosocial decision-making is autonomy. In short, to perceive a decision as reflecting their human nature attributes, people would have to be in the driver's seat, making the decision themselves. Most definitions of autonomy have notions of free choice and self-determination in common (André et al. 2018; Christman 2020; Deci and Ryan 2000; Ryan and Connell 1989; Wertenbroch et al. 2020). For example, Janiesch et al. (2019) posited: "In general, autonomy describes an entity's or agent's ability to act independently and self-determined" (p. 164). Longstanding research traditions in psychology have established autonomy as a fundamental human need (Christman 2020; Deci and Ryan 2000). For instance, in self-determination theory (Deci and Ryan 1985, 2000), the autonomy a person experiences while engaging in a task is a central driver of the intrinsic motivation for performing that task.

In for-profit microlending decisions, however, such autonomy might be less desired, as people have less to gain from seeing themselves as being good at maximizing profits than from seeing themselves as

being good in terms of possessing human nature attributes—and ultimately as good human beings. At the same time, people have more to lose from having full autonomy (instead of giving up autonomy to their decision support system) in for-profit decisions. As people believe that there are objectively right and wrong choices, selecting the right recipients will allow them to maximize their profits, while selecting others could lead to substantial losses and feelings of regret. In consequence, people might be willing to give up autonomy to others with a higher domain knowledge when they are pursuing a clear, objective goal in their decisions, as, for instance, in for-profit microlending. Yet, giving away autonomy to somebody or something else might undermine the perception that they personally (and thus autonomously) selected the option and thereby ultimately prevent the option from feeling right. Just as building a piece of furniture with one's own hands positively affects how much one likes the furniture (Norton et al. 2012), making a prosocial decision with one's own mind might also positively affect how much one feels connected to the recipient.

2.4 Hypotheses Development

Empathy and autonomy, the two factors whose perceived importance is shaped by the context, and, specifically, by contextual differences in terms of the context´s relevance for self-humanization (as described above), can be easily mapped onto the five categories of human nature attributes proposed in Haslam's (2006) review paper: emotional responsiveness, interpersonal warmth, cognitive openness, agency, and depth. First, without empathy, perceiving what someone else feels is difficult, which closely links empathy with emotional responsiveness and interpersonal warmth. Coldness—the antagonist of interpersonal warmth—delimits itself from empathy. Second, without cognitive openness and agency, decision makers cannot make autonomous decisions—they are preconditions for autonomy. Third, the last category, depth, can again be linked to empathy: Without feeling as others feel, how can one achieve a deep understanding of their situation? It is thus not surprising that empathy is seen as one of the most important ways to prevent and overcome dehumanizing (Halpern and Weinstein 2004). Furthermore, autonomy is often withdrawn when someone is dehumanized by others (Haslam 2006), which emphasizes the importance of autonomy.

Another reason why these two factors, the importance of empathy and the importance of autonomy, are central for understanding the contextual differences in the relevance of human nature attributes is the transcendent, moral, and altruistic motives such contexts activate (Batson 1990; Eisenberg and Miller 1987). Moral decisions are often seen as deeply grounded in emotions (Gray et al. 2017; Haidt 2001) and in empathy in particular (Decety and Cowell 2014; Shaw et al. 1994). Prior research has also emphasized the close relationship between the ability to make moral judgments and autonomy—the ability to freely choose actions (Monroe et al. 2017; Nahmias et al. 2014).

2.4.1 Empathy

The first factor stemming from human nature attributes, empathy, is by its very nature not an objective criterion, as different people can have different empathic reactions to the same potential beneficiary of a prosocial decision (Cuff et al. 2016). As in previous research (Dickert et al. 2011; Pavey et al. 2012), we do not use empathy as a measure of individual differences, capacities, or abilities but rather focus on the context-specific importance people grant to this feeling toward others. People can, in consequence, perceive this feeling as more or less relevant for making decisions—which is why it is particularly important for making prosocial decisions (and generally less important for making for-profit decisions). Of course, this is not to say that empathy does not play any role at all in for-profit microlending decisions. For instance, it might allow decision makers to increase the accuracy of their inferences about the likelihood of paying back their loans (Moss et al. 2015) if they emphatically understand the lenders' motivations and emotions. However, in general, building on the idea that prosocial decisions are particularly relevant for self-humanization, we expect the context of prosocial microlending, compared to the context of for-profit microlending, to render decision makers' feelings of empathy more important for making decisions.

> **H1:** Users on prosocial microlending platforms place a higher importance on their empathy with the loan recipients than users on for-profit microlending platforms do.

We expect that the increased importance of empathy for prosocial microlending decisions (compared to for-profit microlending decisions) will ultimately translate into higher levels of algorithm aversion in these contexts, for three main reasons: First, to the extent that people see their own capacity to feel emotions and specifically to feel empathy with the beneficiary as being relevant to making a decision, they will prefer to receive decision support from other actors who also have this capacity. Users might attribute the capacity to feel emotions to other humans but not to computers, because they might be aware of the fact that computers cannot have feelings (Kahn et al. 2006). Additionally, people tend to seek social or parasocial relationships with the source of advice (Önkal et al. 2009; Prahl and van Swol 2017), which works much better when they are getting advice from a human. People do not want to feel empathy only toward the recipient of their prosocial loan but also toward the advisor who supports them in the decision process. However, as empathy is a part of human nature that cannot be possessed by computers (Castelo et al. 2019; Haslam 2006), their perceived lack of empathy for both loan recipients and the platform's user could drive the user's algorithm aversion (Jussupow et al. 2020).

A second reason for increased algorithm aversion is related to the algorithm aversion's antecedents of *cognitive incompatibility, divergent rationalities* (Burton et al. (2020) and *capability* (Jussupow et al. (2020) that we introduced above. Because empathy is not a capability that is ascribed to computers (Castelo et al. 2019) and yet is perceived to be of great importance for prosocial decision tasks, the user can neither map the algorithm's decision processes onto the task requirements nor fully understand and "translate" them. Relatedly, the importance of empathy (which is hard to quantify) makes other more objective criteria that might allow for computing probabilities and inferring objective rankings become

relatively less relevant. For example, Small et al. (2007) showed that in charity domains, people base their decisions on affective reactions, which are not based on objective criteria (see also Slovic et al. 2006). In sum, decisions based on empathy might appear more unstructured and adapted to a world of uncertainty (where expected value calculations are by definition impossible), which creates a mismatch with algorithmic approaches that are usually based on the optimization of quantified criteria (only possible in a world of risk) (Burton et al. 2020). This mismatch, in turn, leads to algorithm aversion.

Third, research on the transparency of algorithms and the understandability of artificial intelligence (e.g., Rader et al. 2018; Shin and Park 2019) has shown that computer decision aids often appear to people as a black box—people do not understand how and why the decision aid has arrived at its recommendation. Especially in situations where empathy is perceived as important and the users are skeptical whether computers are capable of empathy, they will develop questions about how the algorithm works. For example, they may ask themselves: Does the algorithm include the personal story in its calculation? Would, or could, an algorithm incorporate my personal interests? Because they cannot look into the black box, it becomes difficult, if not impossible, to judge whether an algorithm is capable of taking into account or at least approximating subjective feelings like empathy. If people do not believe that algorithms can incorporate the importance they place on empathy, they will not follow the algorithm's recommendations—they become algorithm averse.

H2: The more important empathy is for users, the higher their algorithm aversion.

2.4.2 Autonomy

The second factor stemming from human nature attributes, autonomy, can be understood in a very general way as the ability to make a decision freely and in a self-determined way. While people in general prefer more autonomy over less autonomy, the importance of autonomy differs across contexts (Deci and Ryan 2000). The distinction between *giving* and *giving in*, as two types of motivation for prosocial behavior (Andreoni et al. 2017; Cain et al. 2014; Dana et al. 2007), further illuminates such differences. *Giving* refers to prosocial behavior in which someone engages with full autonomy, willingly, and in the absence of any situational pressure. *Giving in* refers to reluctant prosocial behavior, in which someone engages, for instance, in response to concerns about reputation or social obligation. When people have the opportunity to avoid a situation in which they would be compelled to give in, they usually take it (Cain et al. 2014). Think, for instance, about a shopping center with two exits, in one of which a homeless person is sitting and begging for money. Research has found that more people choose the other exit, avoiding the situation. At the same time, people voluntarily sign up for fundraisers, volunteer in soup kitchens, and browse on prosocial microlending platforms. A key factor in distinguishing these two types of drivers of prosocial behavior is people's perceived autonomy. People want to freely decide to help others rather than feeling compelled to do so because only a free, autonomous decision is relevant for self-humanization. They want to feel the warm glow *because* they

made the decision. The same prosocial behavior would not feel as fulfilling if people performed it reluctantly, due to situational pressures.

The feeling that one *has to* make a loan in for-profit contexts because the opportunity for profit is too good to miss out on (i.e., situational pressures) is much less psychologically meaningful than freely and autonomously *wanting* to make a loan. Of course, some people enjoy the feeling of mastering the process of selecting highly profitable loans, but making such decisions autonomously does not reflect on their degree of self-humanization—of feeling fully human. Taking these considerations together, we hypothesize:

> **H3:** Users on prosocial microlending platforms place a higher importance on their autonomy while making decisions than users on for-profit microlending platforms do.

This context-dependent desire people have to think and feel that they are making autonomous decisions is easily undermined when algorithms and decision support systems come into play (André et al. 2018; Calvo et al. 2020; Wertenbroch et al. 2020). The mere existence of the "human in the loop" discourse (Parasuraman et al. 2000; Sirajum Munir et al. 2013) underlines this point, although often, humans want more than to merely be in the loop. For instance, autonomy can be undermined by recommendations based on past preferences, which make the opinions and preferences of the individual persons unnaturally stable (André et al. 2018), by withholding alternative information or ill-fitting recommendations (Wertenbroch et al. 2020), and by interactive decision aids that restrict the user by adding constraints to the decision process (Pfeiffer et al. 2014). When people ascribe high importance to autonomy, they might be even more sensitive to those restrictions and more or less subtle influences by algorithms.

Additional evidence for the relationship between autonomy and algorithm aversion comes from Jussupow et al. (2020), although they use the related term *agency*, following Komiak and Benbasat (2006). They compared different types of algorithms, which they term *performative* and *advisory* algorithms. Performative algorithms decide or act completely autonomously, without a human's involvement, leaving to the user only the option to delegate a task to the algorithm or not. Advisory algorithms follow a "human-in-the-loop" approach: the user always makes the final decision (see Bonaccio and Dalal 2006), giving them the autonomy to rely on or to ignore the algorithm's advice. Prior research has indeed found that people are more averse to performative algorithms (Palmeira and Spassova 2015) and that they consider algorithmic advice more carefully when it comes from advisory algorithms (Jussupow et al. 2020), supporting the idea that they prefer to keep their autonomy and dislike losing control to the algorithm (Burton et al. 2020). In sum, there is a natural tension between the user's autonomy and the system's autonomy. If users want to use the algorithms and computerized decision support systems, they have to give at least some autonomy to them.

At the same time, placing a high importance on autonomy does not lead to an aversion to advice in *general* to the same extent that it leads to the more specific aversion to *algorithmic* advice. When it comes to decision support, people face the following conundrum: On the one hand, decision support could help them manage and extract the relevant information. This is especially important when people are pursuing a clear, objective goal by their decisions, such as in for-profit microlending, when they might select recipients with the best-fitting interest rate. On the other hand, decision support systems could undermine the perception that *they* personally (and thus autonomously) have selected the option. Algorithmic advice is especially likely to undermine this perceived autonomy if users see the algorithm as a black box and thus cannot understand how the algorithm's advice was computed or if the assumed process the algorithm is following differs widely from the process that users would follow themselves (for instance, by placing less importance on empathy, as machines are assumed to be incapable of feeling empathy). Formally, we state this hypothesis as follows:

H4: The more important autonomy is for the users, the higher their algorithm aversion.

2.4.3 Human-like Decision Support as a Remedy to Dehumanization-induced Algorithm Aversion

When it comes to human advice, people have a much easier time navigating the conundrum mentioned above and balancing their own autonomy and the autonomy given over to the other agent. Extensive research on advice taking (and often advice rejecting, Logg et al. 2019) has demonstrated how well people in their social environments are attuned to navigating and negotiating this conundrum. When interacting with other humans, they can gain much information through social cues and interpersonal connections (Huang and Lin 2011; Joinson 2001; Moon 2000) without giving up independence and autonomy. When dealing with algorithms and computer systems, they may not only lack many of the social cues that underlie this ability but also lack the confidence in accepting and rejecting advice that comes from their extensive experience with human advice givers. Consequently, letting the user perceive the algorithm more as a human than a machine (Epley et al. 2007), by letting the algorithm have or at least imitate human-like characteristics (e.g., emotions), would be a way out of this conundrum.

As research involving the computers as social actors paradigm (CASA) has already demonstrated, humans often show social responses to computers that are comparable to those they show to humans when interacting with them (Nass et al. 1994; Nass and Moon 2000). Human-like decision support taps into this already-existing perception, building on people's tendency to seek a social or parasocial relationship with the source of advice in certain contexts (Prahl and van Swol 2017). In other words, human-like decision support aims not only at communicating effectively but also at building a form of human connection to the user. By clothing the algorithmic decision support in human likeness, system-designers could ultimately prevent users from self-dehumanization in contexts in which they feel a tension between possessing human nature attributes and using (computerized) decision support systems.

The logic behind our next hypotheses is thus that people with high algorithm aversion do not want to use algorithms in a computer-like fashion. However, when algorithms come across as a more human-like decision support, many obstacles against algorithms may disappear or at least become less noticeable. Thus, we formally state the following:

H5: The higher the users' algorithm aversion, the more they prefer human-like decision support.

We have already alluded to the role of autonomy and empathy in driving people's algorithm aversion and ultimately in driving their preference for human-like decision support. Building on that, we expect the type of platform to have an overall effect on the acceptance of human-like decision support.

H6: Users prefer more human-like (and less computer-like) decision support on prosocial microlending platforms than on for-profit microlending platforms.

Figure 2.1 depicts the theoretical framework that we develop in this section.

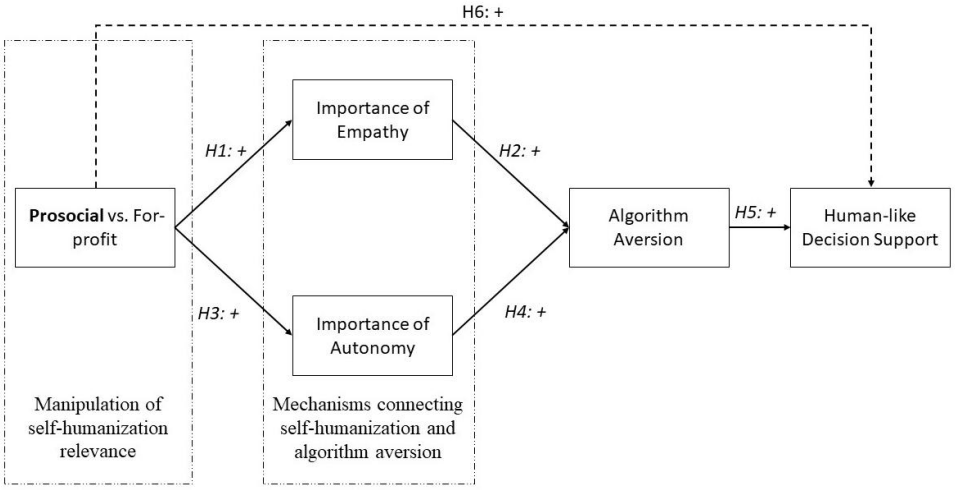


Figure 2.1: Theoretical framework

2.5 Method

2.5.1 Independent Variables and Experimental Design and Procedure

We randomly assigned participants to either a prosocial or a for-profit microlending condition in a between-subjects experimental design (see Appendix A for the experimental stimuli; the appendices are available via http://link.springer.com). Participants in the prosocial (for-profit) condition read an explanation of what prosocial (for-profit) peer-to-peer microlending is and saw three examples of what projects could look like (see Figure 2.3 and Figure 2.4). Next to a project description and an abstract picture, the examples contained information on the loan amount, risk rating, and whether the entrepreneurs had repaid their former loans on time. In the for-profit condition, the only additional information shown was the loan's interest rate.

Because we were not able to derive clear expectations about effect sizes from prior research, we aimed at a sample size that would allow us to detect a small effect. At the same time, we aimed at a sample size large enough to detect simple mediations in order to be able to gain deeper insights into the relationships in the research model. To do so, we relied on the power analysis of Fritz and Mackinnon (2007), which postulated an effect size from 0.14 in standardized betas for small effects. Such an effect size would require a sample size of 462 observations to be detected using a bias-corrected bootstrap with a power of 80%.

We recruited our participants via Amazon Mechanical Turk (MTurk). To ensure that our participants read the provided examples and introduction carefully, we added four comprehension questions at the beginning of the experiment (Goodman et al. 2013). Participants who failed at these questions were automatically excluded. In total, we ended up with 615 US-based participants. Each participant was paid $2 for completing the 15-minute experiment. After we eliminated participants who tried to complete the experiment multiple times (n = 127), who had already participated in a pilot test of this experiment (n = 2), whose origin was not in the US (n = 2), or who made incorrect statements (e.g., an incorrect worker ID (n = 6), 478 participants remained in the final sample (female 40%, male 58%, other 1%, 1% who chose not to provide gender information; mean age = 39.88 with SD: 11.12).

2.5.2 Operationalization of the Dependent Variable

After reading the explanation of the respective experimental conditions, participants answered questions to measure the dependent and control variables. All items can be found in Appendix A.

As a manipulation check for our operationalization of the relevance of self-humanizing (i.e., prosocial versus for-profit context), we used a self-humanization scale based on human nature attributes from the scale of Ruttan and Lucas (2018), at the beginning of the questionnaire. A simple t-test confirms that participants in the prosocial condition indeed found the prosocial context to be more relevant for self-humanization (n = 242, mean = 5.2, SD = .98) than participants in the for-profit condition found the for-profit context (n = 236, mean = 4.4, SD = 1.18, $t(476) = 8.14$, $p = <0.001$, $d = 0.74$).

To measure the importance of empathy, we adopted two scales from the interpersonal reactivity index from Davis (1980, 1983). From the two scales—perspective taking and empathic concern—we created five 7-point Likert items, which we rephrased to fit our focus on the importance of empathy.

To measure the importance of autonomy, we developed a measure based on definitions of autonomy (Christman 2020; Janiesch et al. 2019; Wertenbroch et al. 2020). Additionally, we consulted the need of autonomy scale and the scale for autonomous motivation (Deci and Ryan 2000; Gagné 2003), which we rephrased to fit our focus on the importance of autonomy.

To measure algorithm aversion, we let our participants indicate on a 7-point Likert scale whether they would choose a human supporter or a computerized decision support. The question is one way to

measure algorithm aversion (Jussupow et al. 2020) using the user's choice algorithm aversion measurement and was adopted from Longoni et al. (2019). We also tried to measure algorithm aversion with one of the alternative approaches roughly proposed by Jussupow et al. (2020) that uses the user's evaluation, including items on trust, appropriateness, and authenticity. If the algorithm is evaluated less favorably than the computer on these scales, this would be an indicator of algorithm aversion.

Our scale for human-like decision support is anchored in the theoretical base of dehumanizing and more precise in the human nature attributes. We adapted the items from Ruttan and Lucas (2018) and asked the participants about a list of human nature attributes that a decision support system should be capable of.

As controls, we asked participants about their basic demographics, such as age, gender, and where they currently live. In addition, we added some exploratory questions about the importance of different filters (such as loan amount left, risk rating, etc.).

As discussed above, there are many different causes of algorithm aversion. To rule them out as potential confounds, we added several questions to measure them. First, to measure expectations and expertise, we asked about the frequency with which users had previously used such a microlending platform. Second, to gather information about domain knowledge, we added a single question regarding experience with computerized decision support. Third, we added two scales from Bigman and Gray (2018) about the computer's experiential capability and the computer's capability to think, reason, and plan. Fourth, to measure incentivization through social norms (e.g., information about another user's application of the algorithm; see Burton et al. (2020)), we also added a single question (all questions are listed in Appendix A). We did not control for three additional categories specified by Jussupow et al. (2020) and Burton et al. (2020)—performance, social distance, and human involvement—because they are of little relevance to our context.

For each multi-latent construct, we calculated one standardized factor based on the associated items. For the latent constructs, we examined the convergent and discriminant validity of the measurement instruments. The Cronbach's alphas and composite reliabilities (CR) were greater than the suggested threshold of 0.70, and the values of the average variance extracted (AVE) were above the suggested minimum of 0.50 (see Table 2.3 in Appendix B), except for the importance of autonomy scale. All six items achieved only an AVE of 0.38 and a Cronbach's alpha of 0.67, which suggested a potential issue with the convergent validity. A deeper analysis revealed that two questions loaded poorly on the construct, which was the reason for the low AVE. After removing these two items (3 and 6), we achieved an AVE of 0.49, which is in the acceptable range. Nevertheless, through the removal of these two items, Cronbach's alpha declined (0.64), which was expected because Cronbach's alpha is also driven by the count of items that are combined in the measurement scale. In favor of the higher convergent validity, we decided to base our analyses on the construct with the two removed items, but as a robustness check,

we verified that the results remained robust toward testing the hypotheses with the complete 6-item scale.

To test the discriminant validity, we assessed the factor loadings and cross-loadings (Gefen and Straub 2005). All of the factors loaded higher on the assigned theoretical construct than on any other factor. An additional criterion for establishing discriminant validity demands that the square root of the AVE be larger than any correlation with another construct (Fornell and Larcker 1981). This criterion was also satisfied (see Table 2.3 in Appendix B). We concluded with the HTMT criterion, which is smaller than the threshold of 0.85 (Henseler et al. 2015). In sum, we concluded that our measures exhibited an adequate level of convergent and discriminant validity.

2.6 Results

Table 2.1 summarizes the descriptive statistics, and Table 2.2 along with Figure 2.2, depicts the results of our statistical analyses. As Table 2.1 shows, on the 7-point Likert scale the participants on average rated higher in the prosocial condition than in the for-profit one.

Table 2.1: Descriptive statistics

Variable	**Prosocial condition (N = 242)**		**For-profit condition (N = 236)**	
	Mean	***SD***	***Mean***	***SD***
Relevance of self-humanizing	5.19	0.98	4.39	1.18
Importance of autonomy	5.23	1.00	4.83	1.06
Importance of empathy	4.98	1.13	3.88	1.50
Algorithm aversion	4.52	1.89	4.21	1.96
Preferred human-like decision support	4.40	1.42	3.72	1.49
Age	Mean = 39.88		SD = 11.12	

N = 478

To test our hypotheses H1–H5, we used the seemingly unrelated regression (SUREG) framework, as it allowed us to test our hypotheses while including the control variables in the model and as it is suitable for using binary independent variables.[7] For all analyses, we controlled for age and gender, and while for testing H2 and H4 (influence on algorithm aversion), we also controlled for the already mentioned causes of algorithm aversion: perceived domain knowledge, experience with computerized decision support, incentivization through social norms, and the perceived capability of a computer. Table 2.5 in Appendix B contains more in-depth information on the control variables. The already mentioned Figure 2.2 illustrates our empirical model, with the dotted rectangle marking the SUREG model.

Our results support H1: participants placed a significantly higher importance on their empathy with the loan recipient in the prosocial experimental condition than in the for-profit experimental condition (β =

[7] All calculations were performed with the software STATA/SE 16.1.

0.87; SE = 0.11; p < 0.001). In addition, H2 is supported, which means that a higher importance of empathy leads to higher algorithm aversion (β = 0.28; SE = 0.07; p < 0.001). In other words, participants in the prosocial condition reported a 0.87 higher importance of empathy on a 7-point Likert scale, and with a 1-point increase in this importance, participants reported a 0.28 higher algorithm aversion.

In addition, hypothesis H3 is supported by our results: participants placed a significantly higher importance on their autonomy while making the decision in the prosocial condition than while making the decision in the for-profit condition (β = 0.33; SE=0.09; p<0.001). Furthermore, our model also supports H4, which means that higher importance of autonomy (β = 0.33; SE = 0.08; p < 0.001) leads to higher algorithm aversion.

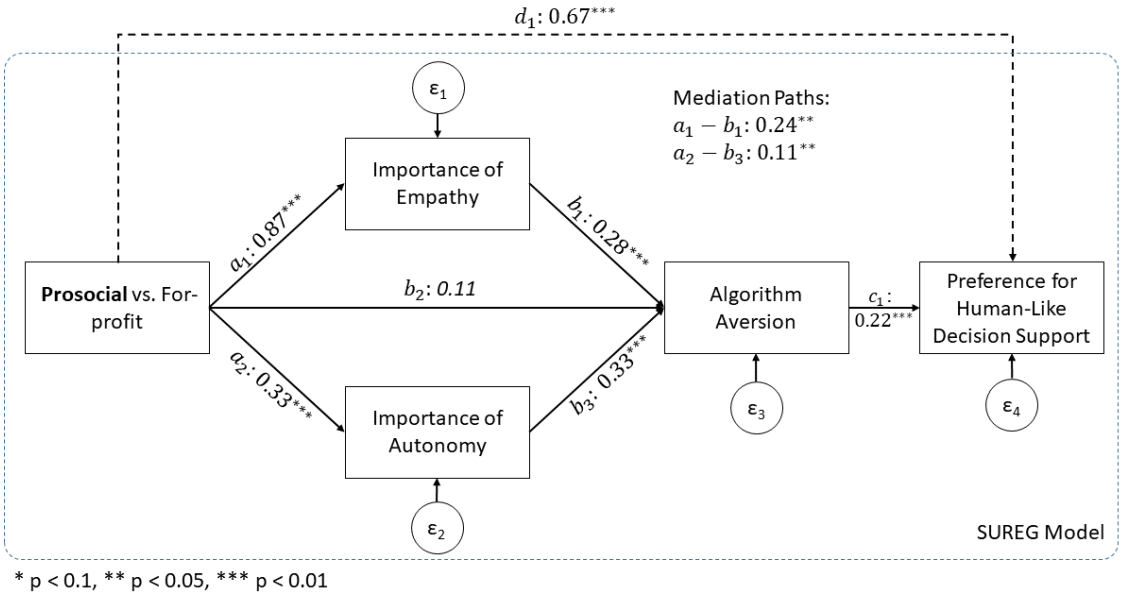


Figure 2.2: Empirical model

To test the relationships postulated in H1 to H4 in detail, we ran a parallel mediation model, allowing the experimental condition to affect algorithm aversion through two mediators, empathy and autonomy. We also included the experimental condition as a direct effect on algorithm aversion in the model (see path b_2 in Figure 2.2) This direct path b2 was not significant (see Table 2.2: β = 0.11; SE = 0.17; p= 0.57). Both indirect paths are significant (95% CI of empathy: [0.11; 0.39] and autonomy [0.04; 0.21]). Because of the significant indirect effects in combination with the non-significant direct effect, our mediation model can be classified as an indirect-only model (Zhao et al. 2010), often also described as full mediation.

Finally, yet importantly, our model also estimates the effect of algorithm aversion on the preference of human-like decision support. As can be seen in Table 2.2, the result is significant (β = 0.22; SE = 0.03; p < 0.001) and positive, supporting H5.

Table 2.2: Empirical results

Hypotheses and path	***β***	***SE***	***P/CI***	***Supported?***
H1 (a_1)	0.87	0.11	<0.001	yes
H2 (b_1)	0.28	0.07	<0.001	yes
H3 (a_2)	0.33	0.09	<0.001	yes
H4 (b_3)	0.33	0.08	<0.001	yes
Indirect effect ($a_1 - b_1$)	0.24	0.07	[0.11; 0.39]	-
Indirect effect ($a_2 - b_3$)	0.11	0.04	[0.04; 0.21]	-
H5 (c_1)	0.22	0.03	<0.001	yes
H6 (d_1)	0.67	0.13	<0.001	yes

Notes: The experimental condition was dummy-coded, with 0 = for-profit and 1 = prosocial. For indirect effects, we used bootstrapped bias-corrected confidence intervals (CI) (with 5,000 resamples), following the recommendation of Preacher and Hayes (2004, 2008).

In order to test our last hypothesis (H6) about the total effect of the experimental condition on human-like decision support, we estimated a simple OLS. As hypothesized, in prosocial decisions (compared to for-profit contexts), human-like support is preferred ($\beta = 0.67$; SE = 0.13; $p < 0.001$). In other words, people reported a 0.67-point higher preference for human-like decision support in the prosocial condition than in the for-profit condition. In summary, our model supports all of our hypotheses.

To explore the robustness of our results, we ran the following three robustness checks, which all followed the same basis specification outlined in Figure 2.2. As a first robustness check, we estimated our model without any control variables (results in Table 2.4 in Appendix B). This led to consistent results with our model reported above, with one meaningful difference: the importance of empathy no longer had a statistically significant effect on algorithm aversion ($\beta = 0.03$; SE = 0.7; p= 0.68, see Table 2.5). In consequence, the indirect effect through empathy was also no longer significant. As algorithm aversion is generally conceived of as a multi-determined phenomenon, not adjusting for other known mechanisms (perceived domain knowledge, experience with computerized decision support, incentivization through social norms, and the perceived capability of a computer) might lead to noisy and biased results (i.e., omitted variables bias). That being said, future research into the intricacies of the relationship between algorithm aversion, the importance of empathy, and the control variables would be needed to illuminate this discrepancy between the different models more thoroughly.

As a second robustness check, we reran our models while additionally including the two omitted items from the importance of autonomy scale. As a third check, we used the alternative algorithm aversion scale based on the evaluation instead of the choice. The robustness checks suggest that our results are robust with regard to these different specifications and the inclusion of these items.

2.7 Discussion

Contexts vary in their affordances for self-humanization. While prosocial contexts are highly relevant for and diagnostic of self-humanization (and human nature attributes in particular), for-profit contexts, by comparison, suppress self-humanization goals. In this paper, we theorize that these differences across contexts lead people to place more importance on empathy and autonomy in prosocial contexts (compared to for-profit contexts) and thereby ultimately induce context-specific algorithm aversion. Human-like decision support holds the promise of remedying this self-humanization-driven algorithm aversion. The results from our experiment lend support to our hypotheses.

First, our experiment shows that decision contexts influence the relevance of self-humanizing (self-humanization was higher in the prosocial than in the for-profit context). The idea that self-humanization is affected by the decision context of digital platforms is, to the best of our knowledge, new and has further implications. Understanding the mechanism of self-humanizing might help us understand what users want from decision support systems and, therefore, how they should be designed. One potential direction for future research would be to broaden the scope of different decision contexts and to investigate the context-specific implications for algorithm aversion and the type of decision support people prefer. The five categories of human nature attributes (emotional responsiveness, interpersonal warmth, cognitive openness, agency, and depth) might carry considerable context-specific implications. For example, there might be domains in which cognitive openness is particularly important for decisions, potentially connecting to research on computational creativity in artificial intelligence (Bentley and Corne 2002; Colton et al. 2012). Moreover, while we concentrated on human nature attributes, uniquely human attributes might also play an interesting role in the design of decision support systems. For instance, they might tighten the connection between humans and algorithms, because these attributes can be shared with machines. In particular, when users consider attributes such as logic and rationality to be important criteria, algorithm aversion might decrease, potentially enabling the acceptance of different kinds of decision support systems.

Second, our experimental results provide further evidence for the idea that empathy is a major factor when it comes to prosocial behavior, and especially that this is also the case for prosocial decisions on digital platforms (H1). We thereby expand on existing research, which has demonstrated the importance of empathy in (non-digital) prosocial behavior (Batson et al. 1987; Davis 2015; Loewenstein and Small 2007; Small and Cryder 2016). Moreover, our results lend support to the idea that autonomy is particularly important for prosocial behavior (H3), which is consistent with prior research—for instance, with the results from Weinstein and Ryan (2010), Gagné (2003), and Pavey et al. (2012). We can conclude that participants want not only to feel empathy for a beneficiary, but also to choose and decide freely in favor of a specific beneficiary.

Third, our experimental results support the proposition that empathy (H2) and autonomy (H4) lead to higher algorithm aversion. We thereby contribute to the burgeoning research stream on the antecedents for algorithm aversion (Burton et al. 2020; Jussupow et al. 2020). In particular, we find that when empathy is seen as an important capability for performing a task, humans as advisors have clear advantages over computers because feeling empathy is a human nature attribute. This finding is obviously related to existing causes of algorithm aversion, such as cognitive incompatibility and divergent rationalities between computers and humans. Furthermore, we find support for the argument that interacting with a human instead of a computer might help users control the process and balance their own autonomy and the autonomy given to the other agent (human or computer).

Fourth, our results provide evidence that algorithm aversion has direct implications for the preferred type of support system. More concretely, a higher algorithm aversion generates the desire to have more human-like decision support, which builds on and connects to the work of Castelo et al. (2019), who already showed that human-likeness could enhance the use of algorithms in more subjective tasks, and to the work of Seeger et al. (2021), who discussed human-like versus computer-like tasks. The question of how to achieve human-like decision support remains highly relevant for the field.

Computerized agents might be seen as missing some human nature attributes, as argued earlier, such as the ability to experience (moral) authenticity (Bigman and Gray 2018; Jago 2019) or empathy. Research on human-computer interaction has already recognized this issue (Picard 2003) and points to potential ways to overcome those deficits—for example, by the use of bio signals like EEG (Song et al. 2020), eye-tracking (Bradley et al. 2008; Pfeiffer et al. 2020), or facial expression (Li and Deng 2020), which allows the system to detect the feelings of the user and thus can take them into account for its suggestions. Another possibility is, as mentioned previously, the anthropomorphization of the decision support system, which is also a new and growing research field. An anthropomorphized conversational agent could emulate human-to-human communication (e.g., using natural language) (Schuetzler et al. 2014). The use of natural language is only one of many possibilities of creating more human-like decision support (Gnewuch et al. 2017). The literature on anthropomorphized conversational agents suggests different cues (Seeger et al. 2021), such as human identity cues (e.g., visual representation (Qiu and Benbasat 2009), verbal cues (e.g., emotional expressions (de Visser et al. 2016)), context-sensitive responses (Knijnenburg and Willemsen 2016)), and non-verbal cues (e.g., emoticons and response delays (Gnewuch et al. 2018)). Yet, even without going to the great length of simulating a complete human conversation, developing a deeper understanding of how self-humanization goals drive people to prefer human-like decision support systems is a fruitful starting point for designing and fine-tuning various sustainable decision support systems.

Finally, we show not only that users in prosocial contexts prefer human-like decision support more strongly than in for-profit contexts, but also that both in prosocial and for-profit contexts, users value human-like decision support (t-tests of the means values with the scale average of 3.5 of the human-like

decision support scale support this finding; prosocial mean = 4.40, SD = 1.42; t(241) = 9.90, p <= 0.001; for-profit mean = 3.72, SD = 1.50; t(235) = 2.30, p =0.01). As described at the very beginning of this paper, human-like decision support is on the rise, and the observation that these support systems are preferred in both types of microlending is thus quite revealing. We have to point out, though, that in both types of microlending, a human is the receiver of the loan. On platforms where humans are not "part" of the choice set from which people choose, for example when the alternatives are share-trading options, we would expect that human-like attributes might be of less importance.

2.8 Contributions, Limitations, and Future Research

Our results have implications for both theory and practice. We contribute to theory by highlighting self-humanizing as an important theoretical lens for understanding contextual differences in algorithm aversion. Contexts differ in their affordances for self-humanization, and the two mechanisms outlines in our framework—the importance of empathy and the importance of autonomy— connect these contextual differences with users' degree of algorithm aversion and, ultimately, their preference for human-like decision support. To the best of our knowledge, autonomy and empathy have not been considered in parallel before in the field of digital platforms, although our results indicate that they should be considered when theorizing about user behavior on prosocial platforms. They complement other factors that have often been studied, such as ease of use, perceived usefulness, and enjoyment (Dwivedi et al. 2015; Gefen and Straub 2000; Pavalou 2003), and future research is needed to investigate the interaction between these factors and empathy as well as autonomy.

Our research also has additional practical implications for designing sustainable decision support systems. At the current stage of technological development, it would be possible to create a conversational agent that is able to fulfill the user's need for empathy and autonomy while lowering algorithm aversion through means like anthropomorphizing, the use of facial expressions, and emotion detection. It is even possible that such a system would not only help the user with a one-time usage of a platform, but also reinforce future prosocial actions (Penner 2002) and thereby increase the overall welfare in the world.

One limitation of the current research is that our experiment did not use actual users of microlending platforms, but MTurkers as participants, although MTurk samples might be more representative as student samples (Chandler et al. 2014). Moreover, several comprehension checks were implemented in an attempt to prevent concerns about speedy and low-effort responses. Another limitation is that the usual caveats of using mediation models for cross-sectional data apply, and we encourage future research to replicate our results to confirm their robustness. An ecologically valid field experiment that moves beyond hypothetical questionnaire responses to consequential lending decisions would be particularly desirable.

Yet another limitation lies in our measurement scale for the importance of autonomy, which could be improved in terms of convergent validity. All items should be analyzed carefully, and a new, more extensive and reliable scale should be developed based on the definition of autonomy. Finally, our robustness checks suggest that the relationship between the importance of empathy and algorithm aversion could be more complicated. Future research should further explore the interplay between empathy, algorithm aversion, and its other antecedents proposed by prior research.

The theoretical framework of self-humanization might provide guidance not only on how to design a decision support system with attributes that are typical for a human (i.e., human-like) but also on other aspects of decision support, for example, the point of time when the support is provided. A decision support system that provides support right at the beginning of a decision process might decrease self-humanization because by restricting user autonomy early-on, it might not leave the user room to fully feel as a human. In contrast, a system that steps in later in the process could give the users the opportunity to fulfill their self-humanization needs first, for example, when the users have had the chance to develop emotional responsiveness to alternatives being decided upon or to create a feeling of interpersonal warmth or agency without being interrupted or undermined by a technical support system. We propose that future research should investigate the influence of the point of time of decision support on self-humanization and its implications for the design of decision support systems.

There is much research on decision support systems, such as recommender and interactive decision aids, but very little on the interplay between those systems and the decision context in which the user is acting. When should we use which type of decision support? How does the decision context affect the relevance of self-humanization? In turn, how important are different factors to users, such as their empathy and autonomy, and how should decision support systems interact with users? As we believe this paper illustrates, much can be gained by bridging multiple research fields and by integrating insights from the psychology into the research field of Information Systems about why and when people act prosocially and lend money. The very existence of algorithm aversion might suggest that many IT artifacts are developed with a focus not on the human but on rather instrumental objectives, such as economic goals. This focus, while often taken for granted and not explicitly acknowledged, can lead to dehumanization (Moore and Piwek 2017). By introducing self-humanization as a theoretical framework, our paper highlights the importance of and facilitates the integration of humanistic values into Information Systems research. We thus contribute to the recently raised call for a stronger sociotechnical perspective in Information Systems (e.g., made by Sarker et al., (2019)).

2.9 References

Adadi A, Berrada M (2018) Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). IEEE Access 6:52138–52160

Aggarwal CC (2016) Recommender systems. Springer, Cham

Allison T, Davis B, Short J, Webb J (2015) Crowdfunding in a prosocial microlending environment: examining the role of intrinsic versus extrinsic cues. Entrep Theory Pract 39:53–73

Allison TH, McKenny AF, Short JC (2013) The effect of entrepreneurial rhetoric on microlending investment: an examination of the warm-glow effect. J Bus Venturing 28:690–707

André Q, Carmon Z, Wertenbroch K, Crum A, Frank D, Goldstein W, Huber J, van Boven L, Weber B, Yang H (2018) Consumer choice and autonomy in the age of artificial intelligence and big data. Cust Need Solut 5:28–37

Andreoni J (1990) Impure altruism and donations to public goods: a theory of warm-glow giving. Econ Theory 100:464

Andreoni J, Rao JM, Trachtman H (2017) Avoiding the ask: a field experiment on altruism, empathy, and charitable giving. J Polit Econ 125:625–653

Bain PG, Kashima Y, Haslam N (2006) Conceptual beliefs about human values and their implications: human nature beliefs predict value importance, value trade-offs, and responses to value-laden rhetoric. J Pers Soc Psychol 91:351–367

Barasch A, Levine EE, Berman JZ, Small DA (2014) Selfish or selfless? On the signal value of emotion in altruistic behavior. J Pers Soc Psychol 107:393–413

Barredo Arrieta A, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, Garcia S, Gil-Lopez S, Molina D, Benjamins R, Chatila R, Herrera F (2020) Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. Inf Fusion 58:82–115

Batson CD (1990) How social an animal? The human capacity for caring. Am Psychol 45:336

Batson CD (2014) The altruism question: Toward a social-psychological answer, 1st edn. Psychology Press, New York Batson CD, Fultz J, Schoenrade PA (1987) Distress and empathy: two qualitatively distinct vicarious emotions with different motivational consequences. J Pers 55:19–39

Bentley PJ, Corne DW (2002) An introduction to creative evolutionary systems. Creative Evo Sys:1–75

Berman JZ, Barasch A, Levine EE, Small DA (2018) Impediments to effective altruism: the role of subjective preferences in charitable giving. Psychol Sci 29:834–844

Bigman YE, Gray K (2018) People are averse to machines making moral decisions. Cognition 181:21–34

Bonaccio S, Dalal RS (2006) Advice taking and decision-making: an integrative literature review, and implications for the organizational sciences. Organ Behav Hum Dec 101:127–151

Bradley MM, Miccoli L, Escrig MA, Lang PJ (2008) The pupil as a measure of emotional arousal and autonomic activation. Psychophysiology 45:602–607

Bruton GD, Khavul S, Chavez H (2011) Microlending in emerging economies: building a new line of inquiry from the ground up. J Int Bus Stud 42:718–739

Burton JW, Stein M-K, Jensen TB (2020) A systematic review of algorithm aversion in augmented decision making. J Behav Decis Making 33:220–239

Cain DM, Dana J, Newman GE (2014) Giving versus giving in. Acad Manag Ann 8:505–533

Calvo RA, Peters D, Vold K, Ryan RM (2020) Supporting human autonomy in AI systems: a framework for ethical enquiry. In: Ethics of Digital Well-Being. Springer, Cham, pp 31–54

Castelo N, Bos MW, Lehmann DR (2019) Task-dependent algorithm aversion. J Mark Res 56:809–825

Caviola L, Schubert S, Nemirow J (2020) The many obstacles to effective giving. Judgm Decis Mak 15:159

Chandler J, Mueller P, Paolacci G (2014) Nonnaïveté among Amazon Mechanical Turk workers: consequences and solutions for behavioral researchers. Behav Res Methods 46:112–130

Christman J (2020) Autonomy in moral and political philosophy. In: Metaphysics Research Lab, Stanford University (ed) The Stanford Encyclopedia of PhilosophyColton S, Wiggins GA, others (2012) Computational creativity: the final frontier? In: Proceedings of the 20th European Conference on Artificial Intelligence, Montpellier, pp 21–26

Cuff BM, Brown SJ, Taylor L, Howat DJ (2016) Empathy: a review of the concept. Emot Rev 8:144–153

Dana J, Weber RA, Kuang JX (2007) Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. Econ Theory 33:67–80

Davis MH (1980) A multidimensional approach to individual differences in empathy. American Psychological Association Washington, DC

Davis MH (1983) Measuring individual differences in empathy: evidence for a multidimensional approach. J Pers Soc Psychol 44:113

Davis MH (2015) Empathy and prosocial behavior. In: Schroeder DA (ed) The Oxford handbook of prosocial behavior. Oxford Univ. Press, Oxford

Dawes RM (1979) The robust beauty of improper linear models in decision making. Am Psychol 34:571–582

Decety J, Cowell JM (2014) The complex relation between morality and empathy. Trends Cogn Sci 18:337–339

Deci EL, Ryan RM (1985) The general causality orientations scale: self-determination in personality. J Res Pers 19:109–134

Deci EL, Ryan RM (2000) The "what" and "why" of goal pursuits: human needs and the self-determination of behavior. Psychol Inq 11:227–268

de Visser EJ, Monfort SS, McKendrick R, Smith MAB, McKnight PE, Krueger F, Parasuraman R (2016) Almost human: anthropomorphism increases trust resilience in cognitive agents. J Exp Psychol 22:331–349

Dickert S, Sagara N, Slovic P (2011) Affective motivations to help others: a two-stage model of donation decisions. J Behav Decis Making 24:361–376

Dietvorst BJ, Simmons JP, Massey C (2015) Algorithm aversion: people erroneously avoid algorithms after seeing them err. J Exp Psychol Gen 144:114–126

Dietvorst BJ, Simmons JP, Massey C (2018) Overcoming algorithm aversion: people will use imperfect algorithms if they can (even slightly) modify them. Manag Sci 64:1155–1170

Dunn EW, Aknin LB, Norton MI (2014) Prosocial spending and happiness. Curr Dir Psychol Sci 23:41–47

Dwivedi YK, Wastell D, Laumer S, Henriksen HZ, Myers MD, Bunker D, Elbanna A, Ravishankar MN, Srivastava SC (2015) Research on information systems failures and successes: status update and future directions. Inf Syst Front 17:143–157

Eisenberg N, Miller PA (1987) The relation of empathy to prosocial and related behaviors. Psychol Bull 101:91–119

Epley N, Waytz A, Cacioppo JT (2007) On seeing human: a three-factor theory of anthropomorphism. Psychol Rev 114:864–886

Fritz MS, Mackinnon DP (2007) Required sample size to detect the mediated effect. Psychol Sci 18:233–239

Gagné M (2003) The role of autonomy support and autonomy orientation in prosocial behavior engagement. Motiv Emot 27:199–223

Galak J, Small D, Stephen AT (2011) Microfinance decision making: a field study of prosocial lending. J Mark Res 48:130-137

Garcia-Garcia JM, Penichet V, Lozano MD (2017) Emotion detection: a technology review. In: Proceedings of the XVIII international conference on human computer interaction, pp 1–8

Gefen D, Straub D (2000) The relative importance of perceived ease of use in IS adoption: a study of e-commerce adoption. J Assoc Inf Syst 1, Article 8

Gnewuch U, Morana S, Maedche A (2017) Towards designing cooperative and social conversational agents for customer service. In: Proceedings of the International Conference on Information Systems, Seoul, South Korea

Gnewuch U, Morana S, Adam M, Maedche A (2018) Faster is Not Always Better: Understanding the Effect of Dynamic Response Delays in Human-Chatbot Interaction. In: European Conference on Information Systems (ECIS2018), Portsmouth, United KingdomGoodman JK, Cryder CE, Cheema A (2013) Data collection in a flat world: the strengths and weaknesses of Mechanical Turk samples. J Behav Decis Making 26:213–224

Gordon C, Leuski A, Benn G, Klassen E, Fast E, Liewer M, Hartholt A, Traum DR (2019) PRIMER: an emotionally aware virtual agent. In: IUI Workshops, Los Angeles, USA

Gray K, Schein C, Cameron CD (2017) How to think about emotion and morality: circles, not arrows. Curr Opin Psychol 17:41–46

Grove WM, Zald DH, Lebow BS, Snitz BE, Nelson C (2000) Clinical versus mechanical prediction: a meta-analysis. Psychol Assessment 12:19–30

Haas P, Blohm I, Leimeister JM (2014) An empirical taxonomy of crowdfunding intermediaries. In: Proceedings of the International Conference on Information Systems - Building a Better World through Information Systems. Association for Information Sytems, AIS Electronic Library (AISeL)Hafenbrädl S, Waeger D, Marewski JN, Gigerenzer G (2016) Applied decision making with fast-and-frugal heuristics. J Appl Res Mem Cogn 5:215–231

Haidt J (2001) The emotional dog and its rational tail: a social intuitionist approach to moral judgment. Psychol Rev 108:814–834

Halpern J, Weinstein HM (2004) Rehumanizing the other: empathy and reconciliation. Hum Rights Q 26:561–583

Hamilton DL, Sherman SJ (1996) Perceiving persons and groups. Psychol Rev 103:336–355

Haslam N, Rothschild L, Ernst D (2000) Essentialist beliefs about social categories. Brit J Soc Psychol 39 (Pt 1):113–127

Haslam N (2006) Dehumanization: an integrative review. Pers Soc Psychol Rev 10:252–264

Haslam N, Bastian B, Bissett M (2004) Essentialist beliefs about personality and their implications. Pers Soc Psychol B 30:1661–1673

Haslam N, Bain P, Douge L, Lee M, Bastian B (2005) More human than you: attributing humanness to self and others. J Pers Soc Psychol 89:937–950

Haslam N, Kashima Y, Loughnan S, Shi J, Suitner C (2008) Subhuman, inhuman, and superhuman: contrasting humans with nonhumans in three cultures. Soc Cogn 26:248–258

Häubl G, Trifts V (2000) Consumer decision making in online shopping environments: the effects of interactive decision aids. Mark Sci 19:4–21

Henseler J, Ringle CM, Sarstedt M (2015) A new criterion for assessing discriminant validity in variance-based structural equation modeling. J Acad Mark Sci 43:115–135

Herzenstein M, Sonenshein S, Dholakia UM (2011) Tell me a good story and i may lend you money: the role of narratives in peer-to-peer lending decisions. J Mark Res 48:138-149

Hoffrage U, Hafenbrädl S, Marewski JN (2018) The fast-and-frugal heuristics program. In: The Routledge international handbook of thinking and reasoning. Routledge, New York, pp 325–345

Huang J-W, Lin C-P (2011) To stick or not to stick: the social response theory in the development of continuance intention from organizational cross-level perspective. Comput Hum Behav 27:1963–1973

Inbar Y, Cone J, Gilovich T (2010) People's intuitions about intuitive insight and intuitive choice. J Pers Soc Psychol 99:232–247

Jago AS (2019) Algorithms and authenticity. Acad Manag Discov 5:38–56

Janiesch C, Fischer M, Winkelmann A, Nentwich V (2019) Specifying autonomy in the Internet of Things: the autonomy model and notation. Inf Syst E-Bus Manag 17:159–194

Joinson AN (2001) Self-disclosure in computer-mediated communication: the role of self-awareness and visual anonymity. Eur J Soc Psychol 31:177–192

Jung D, Dorner V, Glaser F, Morana S (2018) Robo-advisory. Bus Inf Syst Eng 60:81–86

Jussupow E, Benbasat I, Heinzl A (2020) Why are we averse towards algorithms? A comprehensive literature review on algorithm aversion:1–16 In: Frantz Rowe (ed) 28th European Conference on Information Systems - Liberty, Equality, and Fraternity in a Digitizing World, ECIS 2020, Marrakech, Morocco, June 15-17, 2020 : Proceedings. AISeL, Atlanta, GA, pp 1–16

Kahn PH, Ishiguro H, Friedman B, Kanda T (2006) What is a human? Toward psychological benchmarks in the field of human-robot interaction. In: 15th IEEE International Symposium on Robot and Human Interactive Communication. IEEE, pp 364–371

Knijnenburg BP, Willemsen MC (2016) Inferring capabilities of intelligent agents from their external traits. ACM Trans Interact Intell Syst 6:1–25

Komiak, Benbasat (2006) The effects of personalization and familiarity on trust and adoption of recommendation agents. MIS Q 30:941

Lankton N, McKnight DH, Tripp J (2015) Technology, humanness, and trust: rethinking trust in technology. J Assoc Inf Syst 16:880–918

Li S, Deng W (2020) Deep facial expression recognition: a survey. IEEE T Affect Comput.

Loewenstein G, Small DA (2007) The scarecrow and the tin man: the vicissitudes of human sympathy and caring. Rev Gen Psychol 11:112–126

Logg JM, Minson JA, Moore DA (2019) Algorithm appreciation: people prefer algorithmic to human judgment. Organ Behav Hum Dec 151:90–103

Longoni C, Bonezzi A, Morewedge CK (2019) Resistance to medical artificial intelligence. J Consum Res 46:629–650

Maedche A, Legner C, Benlian A, Berger B, Gimpel H, Hess T, Hinz O, Morana S, Söllner M (2019) AI-based digital assistants. Bus Inf Syst Eng 61:535–544

Martinez LMF, Zeelenberg M, Rijsman JB (2011) Behavioural consequences of regret and disappointment in social bargaining games. Cogn Emot 25:351–359

Meehl PE (1954) Clinical versus statistical prediction: a theoretical analysis and a review of the evidence. University of Minnesota Press, Minneapolis

Monroe AE, Brady GL, Malle BF (2017) This isn't the free will worth looking for. Soc Psychol Pers Sci 8:191–199

Moon Y (2000) Intimate exchanges: using computers to elicit self-disclosure from consumers. J Consum Res 26:323–339

Moore P, Piwek L (2017) Regulating wellbeing in the brave new quantified workplace. Empl Relat 39:308–316

Moss TW, Neubaum DO, Meyskens M (2015) The effect of virtuous and entrepreneurial orientations on microfinance lending and repayment: a signaling theory perspective. Entrep Theory Pract 39:27–52

Nahmias E, Shepard J, Reuter S (2014) It's OK if 'my brain made me do it': people's intuitions about free will and neuroscientific prediction. Cogn 133:502–516

Nass C, Moon Y (2000) Machines and mindlessness: social responses to computers. J Soc Issues 56:81–103

Nass C, Steuer J, Tauber ER (1994) Computers are social actors. In: Adelson B (ed) Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, New York

Neth H, Gigerenzer G (2015) Heuristics: tools for an uncertain world. In: Emerging trends in the social and behavioral sciences. Wiley Online Library, pp 1–18

Norton MI, Mochon D, Ariely D (2012) The IKEA effect: when labor leads to love. J Consum Psychol 22:453–460

Önkal D, Goodwin P, Thomson M, Gönül S, Pollock A (2009) The relative influence of advice from human experts and statistical methods on forecast adjustments. J Behav Decis Making 22:390–409

Palmeira M, Spassova G (2015) Consumer reactions to professionals who use decision aids. Eur J Mark 49:302–326

Parasuraman R, Sheridan TB, Wickens CD (2000) A model for types and levels of human interaction with automation. IEEE T Syst Man Cybern A 30:286–297

Pavalou PA (2003) Consumer acceptance of electronic commerce: integrating trust and risk with the technology acceptance model. Int J Electron Comm 7:101–134

Pavey L, Greitemeyer T, Sparks P (2012) "I help because I want to, not because you tell me to": empathy increases autonomously motivated helping. Pers Soc Psychol B 38:681–689

Penner LA (2002) Dispositional and organizational influences on sustained volunteerism: an interactionist perspective. J Soc Issues 58:447–467

Pfeiffer J, Benbasat I, Rothlauf F (2014) Minimally restrictive decision support systems. In: Proceedings of the International Conference on Information Systems, Auckland, New Zealand

Pfeiffer J, Pfeiffer T, Meißner M, Weiß E (2020) Eye-tracking-based classification of information search behavior using machine learning: evidence from experiments in physical shops and virtual reality shopping environments. Inf Syst Res 31:675–691

Picard RW (2003) Affective computing: challenges. Int J Hum-Comp St 59:55–64

Prahl A, van Swol L (2017) Understanding algorithm aversion: when is advice from automation discounted? J Forecast 36:691–702

Qiu L, Benbasat I (2009) Evaluating anthropomorphic product recommendation agents: a social relationship perspective to designing information systems. J Manag Inform Syst 25:145–182

Rader E, Cotter K, Cho J (2018) Explanations as mechanisms for supporting algorithmic transparency. In: Mandryk R, Hancock M (eds) Engage with CHI. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Montréal. ACM, New York

Ruttan RL, Lucas BJ (2018) Cogs in the machine: the prioritization of money and self-dehumanization. Organ Behav Hum Dec 149:47–58

Ryan RM, Connell JP (1989) Perceived locus of causality and internalization: examining reasons for acting in two domains. J Pers Soc Psychol 57:749–761

Sargeant A, Ford JB, West DC (2006) Perceptual determinants of nonprofit giving behavior. J Bus Res 59:155–165

Sarker S, Chatterjee S, Xiao X, Elbanna A (2019) The sociotechnical axis of cohesion for the IS discipline: its historical legacy and its continued relevance. MIS Q 43:695–719

Schuetzler RM, Grimes M, Giboney JS, Buckman J (2014) Facilitating natural conversational agent interactions: lessons from a deception experiment. In: Proceedings of the International Conference on Information Systems, Auckland, New Zealand

Schwartz S (2013) Value priorities and behavior: applying. In: The psychology of values: The Ontario symposium, vol 8

Schwartz SH (1992) Universals in the content and structure of values: theoretical advances and empirical tests in 20 countries. In: Advances in Experimental Social Psychology, vol 25. Elsevier

Seeger A-M, Pfeiffer J, Heinzl A (2021) Texting with human-like conversational agents: designing for anthropomorphism. JAIS:931-967

Shaw LL, Batson CD, Todd RM (1994) Empathy avoidance: forestalling feeling for another in order to escape the motivational consequences. J Pers Soc Psychol 67:879–887

Shin D, Park YJ (2019) Role of fairness, accountability, and transparency in algorithmic affordance. Comput Hum Behav 98:277–284

Sinha R, Swearingen K (2001) Comparing recommendations made by online systems and friends In: Alan F. Smeaton, Jamie Callan (eds) Proceedings of the Second DELOS Network of Excellence Workshop on Personalisation and Recommender Systems in Digital Libraries, DELOS. ERCIM, Dublin, Ireland

Sirajum Munir, John A. Stankovic, Chieh-Jan Mike Liang, Shan Lin (2013) Cyber physical system challenges for human-in-the-loop control. In: 8th International Workshop on Feedback Computing

Slovic P, Finucane ML, Peters E, Macgregor DG (2006) The affect heuristic. In: Slovic P, Lichtenstein S (eds) The construction of preference. Cambridge University Press, Cambridge, pp 434–453

Small DA, Cryder C (2016) Prosocial consumer behavior. Curr Opin Psychol 10:107–111

Small DA, Loewenstein G, Slovic P (2007) Sympathy and callousness: the impact of deliberative thought on donations to identifiable and statistical victims. Organ Behav Hum Dec 102:143–153

Song T, Zheng W, Song P, Cui Z (2020) EEG Emotion recognition using dynamical graph convolutional neural networks. IEEE T Affect Comput 11:532–541

Swangnetr M, Kaber DB (2013) Emotional state classification in patient–robot interaction using wavelet analysis and statistics-based feature selection. IEEE T Hum-Mach Syst 43:63–75

Validi S, Bhattacharya A, Byrne PJ (2015) A solution method for a two-layer sustainable supply chain distribution model. Comput Oper Res 54:204–217

Weinstein N, Ryan RM (2010) When helping helps: autonomous motivation for prosocial behavior and its influence on well-being for the helper and recipient. J Pers Soc Psychol 98:222–244

Wertenbroch K, Schrift RY, Alba JW, Barasch A, Bhattacharjee A, Giesler M, Knobe J, Lehmann DR, Matz S, Nave G, Parker JR, Puntoni S, Zheng Y, Zwebner Y (2020) Autonomy in consumer choice. Mark Lett :429–439

Yadollahi A, Shahraki AG, Zaiane OR (2017) Current state of text sentiment analysis from opinion to emotion mining. ACM Comput Surv 50:1–33

Yeomans M, Shah A, Mullainathan S, Kleinberg J (2019) Making sense of recommendations. J Behav Decis Making 32:403–414

Zhao X, Lynch JG, Chen Q (2010) Reconsidering Baron and Kenny: myths and truths about mediation analysis. J Consum Res 37:197–206

Zheng W-L, Lu B-L (2015) Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. IEEE T Auton Ment Dev 7:162–175

2.10 Supplemental Material

2.10.1 Appendix A: Material

Introduction questionnaire

What is microlending?

When people or businesses need money, they go to the bank and ask for a loan (credit). Microlending is different from normal lending: First, the loan amounts are much smaller (hence “micro"). Second, the entrepreneur(s) typically neither have a well-paying job, nor a good credit history or expensive objects they can use as a guarantee. As in traditional lending, also in microlending there is a risk that the entrepreneurs do not pay back their loan, and thus lenders could lose the money that they have lent.

What is peer-to-peer microlending?

New internet platforms were created on which individual people instead of banks can give microloans to others. Often, these platforms allow the loans to get split up into smaller amounts, so that multiple lenders can contribute a small part to the full loan one entrepreneur will receive.

[***Text only for the prosocial experimental condition******]**

What is prosocial peer-to-peer microlending?

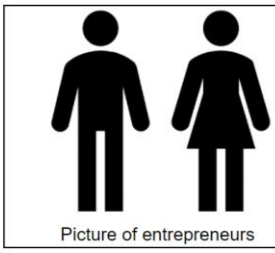
Prosocial means that people who lend money to others are motivated to help them. Consequently, they do not charge interest rates from the entrepreneur(s), and thus do not make a profit from their loan. Rather they take a risk of losing their money if the entrepreneurs are not paying back their loan.

[***Text only for the for-profit experimental condition******]**

What is for-profit peer-to-peer microlending?

For-profit means that people who lend money to others are motivated to gain money through an interest rate. The interest rate normally consists of two parts. The first part is a risk premium, that rises with the risk that an entrepreneur could not pay back the loan - compensating the lender for the risk of losing their money in this case. The second part is a profit for the lender.

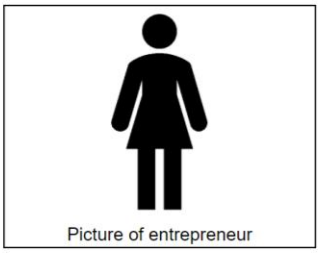
Here are three examples how requests from a for-profit microlending platform could look like:



Picture of entrepreneurs

Chris and Nicole recently started a small business in Senegal to design and create acrylic jewelry and accessories. Initially, they used their savings to buy necessary materials. But now, they need their first funding to keep on producing more jewelry. Increasing their assortment would allow them to sell not just in their small shop, but also to travel to several markets in the region, aiming to keep their business alive and make it sustainable.

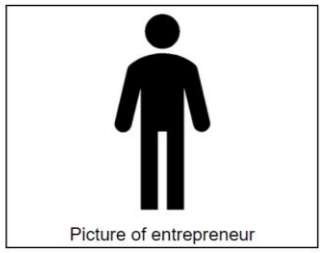
Loan Amount[1]	5,000$
Risk Rating[2]	4.5/5
Repaid on time[3]	Yes
Interest rate[4]	5%



Picture of entrepreneur

Ann is the owner and operator of a small embroidery business in Guatemala, focusing on dresses. Founded four years ago, the business is helping the 30-year-old woman and mother of two children to contribute to her family income. To grow her business and to avoid losing time due to delays in material delivery, she wants to buy thread, fancy cloth, sequins, and pearls of various colors in advance, to have them in stock. The loan would enable her to buy the desired items, as she does not have sufficient cash on hand.

Loan Amount[1]	5,000$
Risk Rating[2]	5/5
Repaid on time[3]	Yes
Interest rate[4]	6%



Picture of entrepreneur

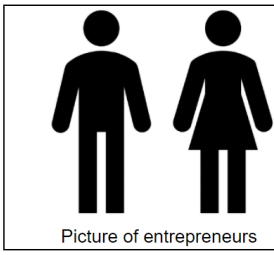
Bob recently started his education as an architect in Bolivia. To support himself during his studies, he works as a tour guide during the weekends. Due the pandemic and the ensuing decline in tourism, he lost his job and cannot afford the monthly payments for his education and flat. To keep his head above water during this challenging period of his life, he asks for an education loan that he aims to pay back after his studies, once he has found a job as an architect.

Loan Amount[1]	5,000$
Risk Rating[2]	3.5/5
Repaid on time[3]	Yes
Interest rate[4]	4%

1. Loan Amount: Describes the amount of the loan that was requested.
2. Risk rating: This key figure is calculated by the peer-to-peer platform and describes how likely the entrepreneur will pay back the loan. A higher value stands for a lower probability of default.
3. Repaid on time: This describes if an entrepreneur has already borrowed a loan on the platform and whether or not they were able to pay it back. If the entrepreneur has never borrowed money before, this information is not available.
4. Interest rate: The interest rate is determined on the base of multiple information by the peer-to-peer platform. A higher risk leads to an increasing interest rate.

Figure 2.3: Examples out of the introduction (prosocial experimental condition)

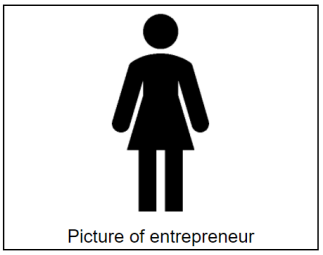
Here are three examples how requests from a prosocial microlending platform could look like:



Picture of entrepreneurs

Chris and Nicole recently started a small business in Senegal to design and create acrylic jewelry and accessories. Initially, they used their savings to buy necessary materials. But now, they need their first funding to keep on producing more jewelry. Increasing their assortment would allow them to sell not just in their small shop, but also to travel to several markets in the region, aiming to keep their business alive and make it sustainable.

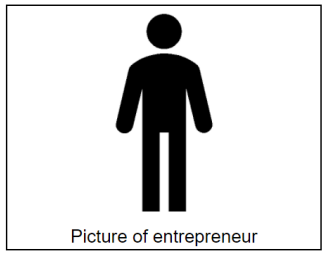
Loan Amount[1]	5,000$
Risk Rating[2]	4.5/5
Repaid on time[3]	Yes



Picture of entrepreneur

Ann is the owner and operator of a small embroidery business in Guatemala, focusing on dresses. Founded four years ago, the business is helping the 30-year-old woman and mother of two children to contribute to her family income. To grow her business and to avoid losing time due to delays in material delivery, she wants to buy thread, fancy cloth, sequins, and pearls of various colors in advance, to have them in stock. The loan would enable her to buy the desired items, as she does not have sufficient cash on hand.

Loan Amount[1]	5,000$
Risk Rating[2]	5/5
Repaid on time[3]	Yes



Picture of entrepreneur

Bob recently started his education as an architect in Bolivia. To support himself during his studies, he works as a tour guide during the weekends. Due the pandemic and the ensuing decline in tourism, he lost his job and cannot afford the monthly payments for his education and flat. To keep his head above water during this challenging period of his life, he asks for an education loan that he aims to pay back after his studies, once he has found a job as an architect.

Loan Amount[1]	5,000$
Risk Rating[2]	3.5/5
Repaid on time[3]	Yes

1. Loan Amount: Describes the amout of the loan the entrepreneur requested.
2. Risk rating: This key figure is calculated by the peer-to-peer platform and describes how likely the entrepreneur will pay back the loan. A higher value stands for a lower probability of default.
3. Repaid on time: This describes if an entrepreneur has already borrowed a loan on the platform and whether or not they were able to pay it back. If the entrepreneur has never borrowed money before, this information is not available.

Figure 2.4: Examples out of the introduction (for-profit experimental condition)

Questions

Manipulation check

When making a microlending decision, I want to feel like I am...

- ... emotional, like I am responsive and warm.
- ... robotic, like I am mechanical and focusing on the hard facts.
- ... superficial, like I have no deep thoughts about entrepreneur(s).
- ... open-minded, like I am receptive for arguments and ideas.
- ... close to the entrepreneur(s).

Importance of autonomy (Adapted from Deci and Ryan 2000; Gagné 2003), Cronbach's α = 0.67 (7-point Likert scale from (1) not important at all (7) extremely important)

If you had to make a decision now for an entrepreneur or a group of entrepreneurs, to what extent would it be important to you ...

- ...to make it without being influenced by others (friends, experts, family, etc.).
- ...to make it without being influenced by features of the website (recommendation systems, chatbots, etc.)
- ...to freely choose from a set of possible options.
- ...to choose an entrepreneur who fits my ideas and opinions.
- ...to choose an entrepreneur who reflects my personal tastes or values.
- ...to be in control of the decision-making process.

Importance of empathy (Adapted from Davis 1980, 1983), Cronbach's α = 0.88 (7-point Likert scale from (1) not important at all (7) extremely important)

If you had to make a decision now, to what extent would it be important to you to decide...

- ...to choose an entrepreneur/entrepreneurs for whom I feel sympathy.
- ...to imagine the situation of the entrepreneur/entrepreneurs.
- ...to feel sorry for the entrepreneur/entrepreneurs.
- ...to feel close to the entrepreneur/entrepreneurs.
- ...to feel concern for the entrepreneur/entrepreneurs.

Algorithm aversion (Adapted from Longoni et al. 2019) (7-point Likert scale from (1) Definitely human supporter (7) Definitely computerized decision support)

On the platform, you can choose between two decision support options: a human who supports you or a computerized decision support system which supports you. Both will first ask you for your preferences and then support you in your decision.

If you had to make a decision now, which support option would you choose to help you with your decision?

Algorithm aversion based on the three evaluation criteria by Jussupow et al. (2020) and the scale by Jago (2019), Cronbach's α = 0.88 (7-point Likert scale from (1) not at all (7) very much so)

Indicate your preference on the provided scale from "not at all" to "very much so".

- To what extent do you trust a human to support you in your decision?
- To what extent do you trust a computer to support you in your decision?
- How appropriate would you find getting help from a human for making this microlending decision?
- How appropriate would you find getting help from a computer for making this microlending decision?
- To what extent do you expect the decision support of a human to be authentic?
- To what extent do you expect the decision support of a computer to be authentic?

Human-like decision support (**Adapted from Ruttan and Lucas 2018), Cronbach's α = 0.87** (7-point Likert scale from (1) not at all (7) very much so)

Imagine that you selected the computerized decision support system. Now you can finetune some of the decision support system's characteristics.

The support system should...

- ... show warmth towards the entrepreneur(s).
- ... be open-minded, i.e. being receptive to ideas and arguments beyond the hard facts about the entrepreneur(s).
- ... be emotional, i.e. it is responsive and warm towards the entrepreneur(s).
- ... be superficial, i.e. having no deep thoughts about the entrepreneur(s).
- ... behave like a computer and not like a human.
- ... be a cold mechanical robot, mathematically optimizing the selection of the entrepreneur(s).

Control Variables:

Causes: Domain experience, experience with computerized decision support, incentivization through social norms

- How frequently have you used such a microlending platform before?
- How frequently were your decisions in this domain supported by a computerized support in the past?

- How many people do you know who are using computerized decision support systems on microlending platforms?

Capability of algorithm/computer (from Bigman and Gray 2018)

To what extend do you think a computer ...

Agency Cronbach's α = 0.91

- ... can communicate with others.
- ... is able of thinking.
- ... can plan its actions.
- ... is intelligent.
- ... has foresight.
- ... is able to think things through.

Experience Cronbach's α = 0.98

- ... is sensitive to pain.
- ... can experience happiness.
- ... can experience fear.
- ... can experience compassion.
- ... can experience empathy.
- ... can experience guilt.

2.10.2 Appendix B: Statistical analyses

Table 2.3: Convergent and discriminant validity

Latent Construct	Cronbach's α	CR	AVE	1	2	3	4	5
1. Human-like decision support	0.8694	0.9013	0.6093	**0.7806**[a]				
2. Importance of autonomy	0.6447	0.7905	0.4855	0.2025***	**0.6968**[a]			
3. Importance of empathy	0.8815	0.9885	0.6049	0.4729***	0.4450***	**0.7778**[a]		
4. Perceived agency capability	0.9108	0.9305	0.6925	0.0968**	0.1724***	0.3293***	**0.8322**[a]	
5. Perceived experience capability	0.9800	0.9838	0.9103	0.0899**	0.2229***	0.4223***	0.4932***	**0.9541**[a]

a The square root of the AVE is shown in the diagonal. The lower triangle shows the correlations between the constructs.
*p < 0.1, ** p < 0.05, *** p < 0.01

Table 2.4: Robustness check without control variables

Hypotheses and path	***β***	***SE***	***P/CI***	***Supported?***
H1 (a_1)	0.86	0.11	<0.001	yes
H2 (b_1)	0.03	0.07	0.680	no
H3 (a_2)	0.33	0.08	<0.001	yes
H4 (b_3)	0.26	0.09	0.005	yes
Indirect effect ($a_1 - b_1$)	0.03	0.07	[-0.11; 0.16]	-
Indirect effect ($a_2 - b_3$)	0.09	0.04	[0.03; 0.18]	-
H5 (c_1)	0.22	0.03	<0.001	yes
H6 (d_1)	0.68	0.13	<0.001	yes

Notes: The experimental condition was dummy-coded, with 0 = for-profit and 1 = prosocial. For indirect effects, we used bootstrapped bias-corrected confidence intervals (with 5,000 resamples), following the recommendation of Preacher and Hayes (2004, 2008).

Table 2.5: Dependent variable: algorithm aversion; with reported betas

DV: Algorithm aversion	**Without controls**	**With controls**
Importance of autonomy	0.26***	0.33***
Importance of empathy	0.03	0.28***
Algorithm control variables		
Domain knowledge		0.06
Experience		-0.19**
Incentivization		-0.05
Perceived agency capability		-0.25***
Perceived experience capability		-0.25*
General control variables		
Gender		
Female		Base
Male		-0.40**
Other		0.18
Do not want to specify		2.07**
Age		-0.01

Notes: * p < 0.1, ** p < 0.05, *** p < 0.01
Partial output of the complete SUREG Model.

3 Paper B: Conversational Agents with Voice

How Social Presence Influences the User Behavior in Microlending Decisions

Pascal Oliver Heßler • Jella Pfeiffer • Matthias Unfried

Abstract

This paper investigates how social cues, used to anthropomorphize assistance systems (ASs), like conversational agents, influences the behavior of users. We find that anthropomorphizing, in our case by giving them a voice, increases social presence, and, in turn, empathy and trust. Yet, we argue that social presence also has negative impacts. We propose—and empirically proof—that a higher social presence also leads to a stronger feeling of being observed. This, in contrary to our hypothesis, results in lower investments. In addition, we found a trend of higher enjoyment with increasing feeling of being observed. This result emphasis the complexity of creating and designing ASs. Through the invocation of social presence anthropomorphizing might have negative and positive effects on outcome variables.

3.1 Introduction

Assistance systems (ASs) are frequently used in online environments, at the same time there are getting smarter and often are more anthropomorphized. This means that users attribute human qualities to ASs, including consciousness, intentions, and emotions, to nonhuman agents (Guthrie, 1993). One reason for anthropomorphizing ASs is that research showed that users tend to perceive those systems as more moral and trustworthy (Epley et al., 2007). Many researchers try to find and test different social cues, in order to achieve an anthropomorphized AS, e.g., for conversational agents (CAs), described as software with which humans interact through natural language. While first researchers start to question the assumption that more social cues lead to more anthropomorphism (Seeger et al., 2021), more studies are needed which study whether increasing anthropomorphism is appropriate for all contexts. Gambino et al. (2020) do emphasize the importance of the context as they showed that different contexts can lead to different perceptions of social cues. Besides the context, research also start to question if social presence, invoked by anthropomorphism, always leads to positive outcomes, like more trust, described as beliefs that the other party has suitable attributes for performing as expected in a specific situation (e.g., Qiu and Benbasat, 2009). Rhim et al. (2022) for example showed that a higher social presence, which connects directly to the perceived anthropomorphism of an AS (Nass and Moon, 2000), leads to negative outcomes like social desirability.

In this paper, we want to understand how different anthropomorphized ASs effect the user behavior. In order to understand context-specific differences as well as potential negative influences of social presence, we decided to use the domain of microlending. Microlending is a financial instrument allowing to request small loans and is mainly used if traditional sources of financing may be unobtainable for them, for instance, because of their lack of collateral (Bruton et al., 2011; Allison et al., 2013). We chose

this domain for our experiment because of three different reasons. First, striving from the idea that people want to see themselves, and to be seen by others, as fully human (Haslam et al., 2005), microlending decisions give lenders the opportunity to highlight their human nature attributes, as helping other have humanization at their front and center (Heßler et al., 2022). Since machines are usually not attributed with human nature attributes (Haslam, 2006), this highlights the necessity to design potential AS in such a way they are perceived as anthropomorphized (more human-like), allowing them to mimic human nature attributes. Second, within the context of microlending, there are two different kinds: the one where users receive interest (for-profit) and the ones where they do not (pro-social). Findings in this context where both the for-profit as well as the pro-social alternatives are available might therefore be generalizable to different settings (more investment like, like the for-profit alternative) and more lending-like (like the pro-social alternatives). Third, as for-profit and pro-social are counterparts regarding the user motives (Heßler et al., 2022) it allows us, in an explorative way, to study if an anthropomorphized AS does lead to more pro-social behavior.

To investigate the potential divergent effects of anthropomorphized ASs through social presence, we designed three different ASs that differ with respect to how anthropomorphized they are. Most studies focus on simple to implement cues like a name (human identity cue) or usage of emoji's (non-verbal cue). While these cues are important and studies could show that they do increase anthropomorphism (Feine et al., 2019), we wanted to shed light on a cue that is less studied and is highly relevant in our context: voice. Voice conveys a variety of information about the communication partner, such as gender and emotions (Schuller et al., 2013; Stern et al., 2021; Scherer, 1995), and could greatly enhance perceived anthropomorphism. In addition, CAs with voice are now widely used, but have only been studied in some context, for example: retail (Poushneh, 2021) and smart speakers (Bentley et al., 2018), but not in more emotional and social contexts like microlending where humans decide about giving money to other humans. Microlending highlights human-nature attributes, and we state that voice might amplify this, as it includes many human nature attributes like emotions.

Our, to different degrees of anthropomorphism, manipulated ASs should result in different levels of social presence. In the context of microlending, we postulate that social presence not only increases trust, as already shown in research, but also the importance of empathy, which is an important emotion in microlending (Davis, 2015). Besides those positive effects, we suspect that social presence does also have negative effects, which are underrepresented in the current state of research. In particular, we expect that social presence can invoke a negative feeling of being observed. We argue that this makes users feel pressure to uphold a positive self-image which might negatively influence service satisfaction and enjoyment but might increase the height of the investment. In order to investigate our research model, we performed an online between-subject experiment, with three different ASs. The results partly support our model and thus the hypothesized causal relationship between factors that have rarely been studied.

We make four main contributions. First, voice does not have such a strong effect on perceived anthropomorphism as expected, as we do not find significant differences to the text-based version of the CA. This emphasis the work from Seeger et al. (2021), who showed that the connection between social cues and anthropomorphism is not always as straightforward as one might think. Developers should not focus only on one specific cue, like the voice, as the effect might be too low. Second, we show that social presence does invoke trust, empathy and the feeling of being observed. From a research perspective, the complex nature of a higher social presence still is of high interest and needs to be researched from new perspectives that fit the respective context, like our focus on feelings like empathy and being observed that we introduced in our experiment. Third, feeling observed does not behave as we expected. Our results show that it does not negatively influence service satisfaction or enjoyment but it decreases the height of investment. This highlights a knowledge gap in literature regarding negative effects of social presence. Fourth, users looked on more different projects in detail when making the investment decision, chose more pro-socially afterwards. This might indicate that pro-social microlending platforms should keep their users' interest in exploring the different projects.

3.2 Theory & Hypothesis

According to the original CASA paradigm, users react socially to a computer exhibiting human-like behavior, mindlessly using social script from human-to-human communication, although they are aware that they are interacting with a computer (Reeves and Nass, 1996; Nass and Moon, 2000). Recently, the reworked approach of CASA (Gambino et al., 2020) considers that users, through increased interaction between anthropomorphized computers and users, developed human-to-media scripts. The nuanced framework suggests a context-dependency: a user's perceptions and reactions to the technology can not only differ from ordinary human-to-human interaction but also from context to context, even in presence of identical social cues (Nass and Moon, 2000; Waytz et al., 2010b). For example, Seeger et al. (2021) could show that social cues did lead to different decrease in perceived anthropomorphism depending on different context, e.g., smart home and healthcare. This approach encourages to look more closely at specific contextual differences.

3.2.1 Impact of Social Cues on User's Perception of CAs

We define anthropomorphism as the pure assignment of human characteristics to non-human objects (Guthrie, 1993). Following the work by Haslam's (2006), we distinguish two different types of human characteristics. First, uniquely human attributes, which can be possessed by machines (e.g., cognitive capabilities, logic) and second, human nature attributes which only can be possessed by humans (e.g., emotional responsiveness, interpersonal warmth). We argue that the assignment of whether an attribute is one or the other type is dynamic due to the increasing intelligence of machines, for example by equipping them with sensors that can detect emotions (Swangnetr and Kaber, 2013). Regarding CAs

that are designed to be human-like interaction partners, it is therefore exciting for us to research the possibility to assign attributes to CAs that, traditionally, are thought to be human nature attributes.

In recent years, research focus has been on providing CAs with so-called social cues (for an overview see Feine et al., 2019). Social cues are defined as design feature, which is salient to the user and triggers a social reaction of the user towards the CA (Nass and Moon, 2000; Fogg, 2002). Social cues have been separated into human identity, verbal and nonverbal cues (Seeger et al., 2021). Seeger et al. (2021) showed that many social cues increase anthropomorphism. However, this relation is not as straightforward as one might think. They found that using only one of the three types of cues does not increase anthropomorphism and using nonverbal cues alone even leads to decreasing anthropomorphism, but when using certain combinations, anthropomorphism is increased. Therefore, research on the effect of social cues is important. We decided to focus on a social cue that has (i) received little focus in research (probably because of a larger implementation effort compared to implementing emoji or simple verbal cues) and (ii) could help transporting particularly human nature attributes.

In combination with several cues-filtered-out theories (e.g., media richness theory, social presence theory) computer mediated communication is by its very nature less informative than face-to-face communication (van Kleef, 2010; Knapp and Daly, 2011), because through limitations of the medium, cues get filtered out. For example, a person's voice includes many hidden cues about the person itself, e.g., where the person originates, the gender, as well as emotional stability and many more (Aung and Puts, 2020; Stern et al., 2021). In addition, it allows the listener inferences about what people want to say, e.g., making jokes. Furthermore, voice is able to transmit many human nature attributes like emotions (Scherer, 1995; Schuller et al., 2013). Hence, not including the voice reduces the cues in the communication and lowers the media richness. Therefore, we decided to add voice to our assistance system in order to increase the anthropomorphism. In addition, Cho et al. (2019) could already demonstrate that voice does increase human-likeness. Finally, the CASA theorem itself also defines voice as social cue (Nass et al., 1997; Nass and Moon, 2000; Mayer et al., 2003).

H1: Voice will increase the perceived anthropomorphism of an assistance system.

In this paper, we distinguish between anthropomorphism and social presence. Anthropomorphism relates to machine design (e.g., social cues), while social presence describes a response to anthropomorphism, which has been described as the feeling of being with another, whether human or computer-mediated intelligence (Biocca et al., 2003; Nass and Moon, 2000; Reeves and Nass, 1996). Previous studies have shown that adding social cues leads to stronger perceived anthropomorphism, and social presence (Araujo, 2018; Feine et al., 2019). Additionally, richer communication mediums (e.g., more social cues relative to less social cues) also increase social presence (Yoo and Alavi, 2001).

H2: Higher perceived anthropomorphism leads to higher social presence.

3.2.2 Impact of Social Presence on Users

Research to date has shown that social presence mainly has positive effects. For example, Gefen and Straub (2004) showed that it can decrease social uncertainty, while Luhmann (2018) argued that social presence is essential for trust building. Social presence also allows more inferences to be made about an AS (Qiu and Benbasat, 2009). Together with the reduction of uncertainty it allows the user to better assess the AS and thus build higher trust (Cyr et al., 2007; Cyr et al., 2009; Qiu and Benbasat, 2009; Gefen and Straub, 2003). Moreover, it has been shown that social presence can also make individuals more comfortable with disclosing (Schuetzler et al., 2018).

H3: With a higher social presence of the AS, users will develop more trust in the AS.

While trust is an often-investigated consequence of high social presence, because it is relevant for many contexts, other effects of social presence exist, which might be more context-dependent. When it comes to decisions that affect other people's lives and ideas, as it is the case in microlending decisions, empathy is an important emotion that might affect the decision (Davis, 2015). Empathy is defined as the ability to take the emotional perspective of someone else—feeling as others—and includes the feeling of sympathy—feeling for others (Batson, 2014; Cuff et al., 2016; Davis, 1983b; Loewenstein and Small, 2007). Empathy is, in addition, an emotion and as such has been shown to be important in microlending (Barasch et al., 2014; Eisenberg et al., 1988; Herzenstein et al., 2011). While empathy is directly connected to taking the perspective of *someone*, social presence indicates if an AS is perceived as *someone*. Therefore, social presence is of great importance when investigating empathy. Based on Haslam (2006), Heßler et al. (2022) argued that people want to feel as human through valuing human nature attributes. Emphasizing empathy as it links to multiple human nature attributes (emotional responsiveness, interpersonal warmth, and depth).

H4: With a higher social presence of the AS, users will place more importance on their empathy.

When it comes to pro-social decisions, altruistic motives are important (Krebs, 1975). Altruism is defined as a motivational state with the ultimate goal of increasing another person's welfare (Batson, 2014), but is also, in itself, heavily discussed. In his research, Batson (2014) examines whether individuals exhibit true altruism or if their actions are motivated by self-interest and a desire to conform to societal expectations and avoid scrutiny from others (see also Barasch et al., 2014). What seems clear is that people behave differently in such altruistic contexts when they feel they are being watched. For example, humans offer more help when someone else is observing what is happening (Cain et al., 2014). In another experiment, consistent with self-awareness theory (Duval and Wicklund, 1972), Batson et al. (1999) were able to show that a focus on the self leads participants to act more fairly.

However, it is not only possible to trigger self-awareness through a mirror, as in Batson et al. (1999). A study by Sah and Peng (2015) showed that increasing anthropomorphization can lead to higher levels of self-awareness. One reason for this effect is that the presence of a communication partner or an AS

makes user reflect more on the decisions they make. In addition, users could try to uphold a favorable image with the AS. This effect is probably stronger with a higher social presence of the AS (Rhim et al., 2022; Sah and Peng, 2015). For example, Haley and Fessler (2005) and Ekström (2012) were able to show that displaying images of eyes can alter behavior, not necessarily with the goal of acting morally, but to create the perception of behaving morally (Freud, 1930; Jones and Pittman, 1982). Observation or the feeling of being observed—by showing pictures of eyes—triggers the feeling that one is actually responsible for one's actions (Batson et al., 1999). A similar phenomenon exists, called social desirability bias, when subjects answer surveys, where they tend to respond in a way that they think is socially desirable to present themselves as having a favorable image (Holbrook et al., 2003; DeMaio, 1984). While most studies only expect a trait effect and collect it to control for it, others directly manipulate it e.g., picture of eyes, hence manipulating the user's state. We suggest that different contexts or different designs of an AS will lead also to effects on the user's state of feeling observed.

Following these findings and arguments, we hypothesize that a higher social presence will create the impression that the AS is also aware of the user's actions and observing what is happening.

H5: A stronger social presence will lead to a stronger feeling of being observed.

Summarizing, social presence drives trust, empathy and the feeling of being observed for a graphical implication see Figure 3.1.

3.2.3 The Impact on Outcome Variables

In the context of microlending decision, the question arises as to how exactly *importance of empathy*, *trust* and *feeling observed* influence the decision-making outcome. From the microlending platform's perspective three dimensions are important that determine the platform's success: the level of investment decisions, the user's service satisfaction, (Verhagen et al., 2014; Rzepka et al., 2021; Sargeant, 1999) and enjoyment (Pfeiffer et al., 2014). Enjoyment describes the intrinsic motivation the willingness to reuse such a service (Pfeiffer et al., 2014; Moon and Kim, 2001) and is therefore also important for the platform's success. All three dimensions are, of course, also important from the user's perspective.

It is quite clear that people who place more value on empathy will also make a higher investment. This can be derived from the fact that these people can empathize more with the persons and thus better understand the reason and the need behind the question of a loan (Loewenstein and Small, 2007; Batson et al., 1999; Eisenberg and Miller, 1987). At the same time, investors may also feel increased enjoyment, because of supporting another person. This can also be referred to the well-studied warm glow effect (Allison et al., 2015). However, through empathy there is also a negative effect. First, the story of the user might be hard to hear, which can lead to negative feelings. Second, the user must always choose between several projects and can always help only a few (in our case only one project). In summary, due to the lack of conclusive evidence, we posit that the effect on enjoyment can go either way and hence we propose a hypothesis in this regard. We do not expect an effect on service satisfaction.

H6a: The higher the importance users attach to empathy, the more they invest.

H6b1: The higher the importance users attach to empathy, the higher their enjoyment.

H6b2: The higher the importance users attach to empathy, the lower their enjoyment.

Trust is a well-studied phenomenon. Especially in an online-environment, where users do not know each other, and platform providers usually have an information advantage, users always try to assess the trustworthiness of another person or platform (Gefen and Straub, 2003; Meyerson et al., 1996). When a platform seems trustworthy this has in general positive effects, as trust can reduce the uncertainty between both parties (Gefen and Straub, 2003; Luhmann, 2018). McKnight et al. (2002b) also showed that higher trust led to higher levels of intention to purchase, which also is related to the level of investment. Furthermore, trust should have a positive effect on enjoyment and service satisfaction, as trust reduces the uncertainty between the platform and the user, which allows the user to concentrate on the actual projects to invest without thinking about the platform.

H7a: The higher the user's trust, the higher the user's investment.

H7b: The higher the user's trust, the higher the user's enjoyment.

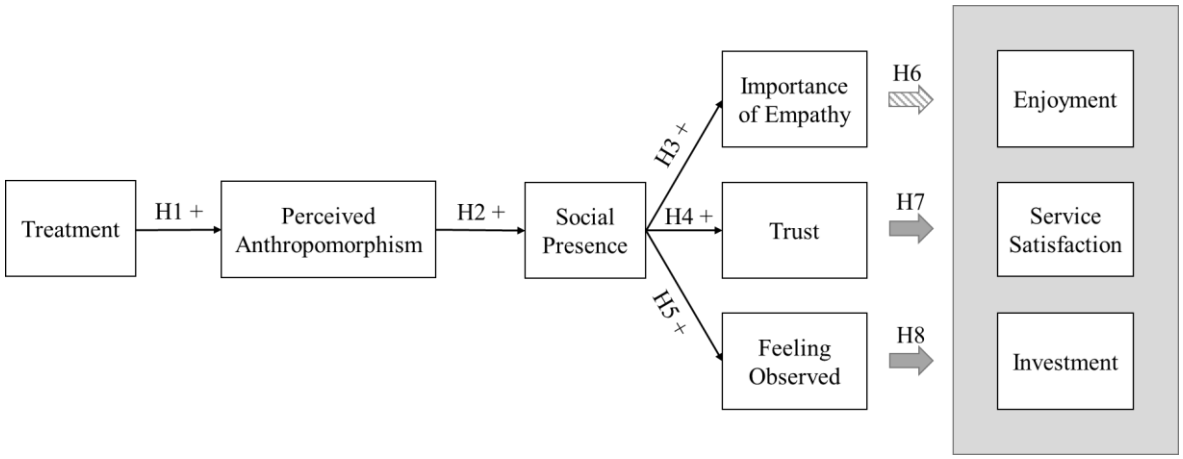
H7c: The higher the user's trust, the higher the user's service satisfaction.

In contrast to trust and empathy, the effect of feeling observed is much less researched. In general, however, this feeling is to be classified as negative (Zwebner and Schrift, 2020). The feeling of being observed, as described above, can trigger self-awareness, which may create pressure on users to present themselves as having a favorable image. As a result of maintaining a positive self-image users may feel forced to invest a higher amount than they actually want to (DeMaio, 1984; Holbrook et al., 2003). This can be linked to the work by Andreoni et al. (2017) on *Giving* and *Giving in*. *Giving* refers to pro-social behavior in which someone engages willingly, and in the absence of any situational pressure. *Giving in* refers to reluctant pro-social behavior, in which someone engages, for instance, in response to concerns about reputation or social obligation. When people have the opportunity to avoid a situation in which they would be compelled to *Give in*, they usually take it (Cain et al., 2014; Weinstein and Ryan, 2010). Think, for instance, at the mentioned experiment from Ekström (2012), where watching eyes did lead to more spending behavior by the supermarket customer (in case that the supermarkets were little frequented). Similar Andreoni et al. (2017) also found that, when people are confronted with a *Giving in* situation, more people try to avoid the situation, e.g., taking another exit. In our scenario, we thus expect lower enjoyment of the investment, with an increased feeling of being watched, as well as lower service satisfaction, since no "way out" is available. Finally, yet importantly, the upholding of a positive self-image also creates a moral expectation, which is why they probably tend to invest more because it is a *Giving in* situation.

H8a: The more users feel observed, the more money they invest.

H8b: The more users feel observed, the less they enjoy the decision.

H8c: The more users feel observed, the less they rate the service satisfaction.



Note: The hatched arrow at H6 only links *importance of empathy* to *enjoyment* and *investment*.
Figure 3.1: Theoretical Model

3.3 Method

3.3.1 Experimental Design & Procedere

We conducted an experiment to test our hypotheses, applying a between-subjects design by manipulating the anthropomorphization of an AS to different degrees. The control treatment was a non-anthropomorphic design that supports choice decisions through filter options (*filter* treatment), such as choosing specific countries. Next, we programmed a state-of-the-art CA in the form of a chatbot (*CA* treatment); using the *botframework* (4.14.2) from Microsoft Azure. As recommended from theory and literature, we implemented for every (sub-)category of social cues (see Table two by Seeger et al. 2021) to achieve an anthropomorphized CA. We used human identity cues (human-like visual representation: image, demographic information: name and gender), verbal cues (social dialog: greetings and dynamic questions, emotional expressions: congratulations, verbal style: self-referencing, context-sensitive responses: the CA included the chosen preferences in its response) and nonverbal cues (emoticons: multiple types, temporal cues: messages were delayed depending on the length of the message, turn-taking gestures: typing indication). Figure 3.2 displays the welcome message of our CA. To investigate whether a voice also has an effect on investment behaviour, the same CA was equipped with a human voice (*CA Voice* treatment). A crucial question when implementing a CA with voice, is whether the voice should be female or male. Findings on the speaker-gender effect are ambiguous. Schild et al. (2019), e.g., found that male voice (with lower pitch) increases trust. Others, however, showed in a learning context, that female voices are preferred and that learners who listened to a female voice, instead of a male voice, did rank higher in test scores (Linek et al., 2010). Lastly most assistance systems, such as Alexa or Siri, use a female voice by default, and users may be used to female voices, so we decided to use a female voice that we have pre-recorded (we have not used text-to-speech to achieve stronger anthropomorphism).

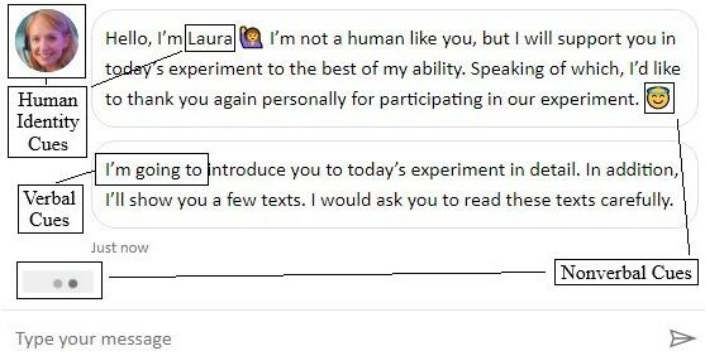


Figure 3.2: Design of Assistance System

Participants were randomly assigned to one of three ASs: filter, CA, or CA Voice. We used Python (3.9) and Django (3.2.7) to build a webpage simulating an investment situation, with a general procedure of five steps (see Figure 3.3). First, identical information about the general procedure of the experiment and about pro-social and for-profit microlending was presented to all participants. Then, the AS asked participants about their preferences for microlending attributes (e.g., social or finance interest, scope of the credit, scope already financed, credit period and region). Next, participants were shown four projects and participants now had the time to take a closer look at the projects and could invest up to €75 in one of them. Finally, we concluded with a questionnaire. The AS sorted the projects based on the participant's preference for social or finance interest.

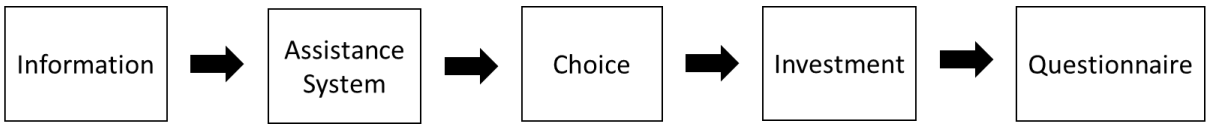


Figure 3.3: General procedure

In total, we had four project templates, which consisted of a general fictional story, a picture, and information about the credit itself, which were manipulated to match the participants' preferences, e.g., the country (for example see Figure 3.4). Additionally, every project exists in two different versions: pro-social and for-profit. The project versions were randomly assigned to participants in order to eliminate the influence of project-specific biases. In other words, the same project was for some of the participants framed as pro-social project and for some as for-profit. The different versions (pro-social versus for-profit) are reflected in multiple changes to the projects. First, a for-profit (pro-social) project had an interest rate of 8% (0%), a fictional "pro-social rating" between 2 and 3 of 10 (7 or 8) as well as an additional paragraph about the motive behind the credit, which was framed in a pro-social or for-profit manner (see second paragraph in Figure 3.4).

To ensure that there is no dominant project and that the for-profit frame of the projects is really perceived as for-profit compared to the respective pro-social frame, we conducted different pre-tests. The pre-tests were conducted in three rounds on prolific, were participants (n_1=80, n_2=80, n_3=80) rated to which degree the projects are pro-social or for-profit. In addition, participants had to choose a project in which

they could imagine investing. We randomly presented four projects to each participant (two in the pro-social and two in the for-profit version). The first two pre-tests revealed a preference for climate-friendly projects, but the final round, after text adaptation, showed a wider range of project choices with expected pro-social and for-profit ratings.

Figure 3.4: Exemplary project in the for-profit layout

A power analysis using G*power (3.1.9.7) was conducted by calculating an ANOVA for three groups. Results indicate a minimum sample size of 252 (f = 0.25, power = 0.95). As the sample sizes for the later-used partial least squares method were smaller, the ANOVA approach was utilized.

3.1.1 Operationalization of the Variables

For measuring perceived anthropomorphism, we used the scale from Waytz et al. (2010b) which was also used by several other studies (Waytz et al., 2014; Seeger et al., 2021). Following previous literature we used the social presence scale from Gefen and Straub (1997, 2000). To proper measure the importance of empathy, we created five 7-point Likert items based on two—perspective taking and empathic concern—of the interpersonal reactivity index from (Davis, 1980, 1983a). This scale was also used and evaluated by Heßler et al. (2022). In order to measure trust, we used the updated trusting beliefs scale from Lankton et al. (2015) which measures trusting beliefs in humans and systems. We adapted Barger and Grandey (2006) service encounter satisfaction scale to fit our context in order to measure service satisfaction. Enjoyment was measured by the scale from Moon and Kim (2001).

To the best of our knowledge, there is no scale for the feeling of being observed. Thus, we developed a new scale. Studies that investigated this topic directly induced the feeling (for example by showing eyes) and did not use it as latent construct. We based our scale on two main attributes of feeling observed: the feeling of being watched and judged by others as argued earlier.

As controls, we asked participants about their basic demographics, such as age, income, and gender. In addition, we added some control variables: Social desirability bias (German version: Winkler et al. 2006

based on Paulhus 1986), dispositional anthropomorphism (Waytz et al., 2010a), dispositional trust technology (McKnight et al., 2002a) and risk aversion (Richter et al., 2013). We included risk because this might in general influence the investment height (Kahneman and Tversky, 2013). An overview about the used scales is included in the Appendix see Table 3.4.

3.3.2 Sample and Model Validation

We recruited participants at the *Justus-Liebig-University*. The experiment was conducted online. We checked with a control question if user could hear a specific audio sound, to ensure people actually could hear the CA in the voice treatment. If they answered wrong, they were excluded. Every participant received €6 for participation and could win €75 with a probability of 8%. If participants won, we invested the amount according to their decision on a real-world microlending platform. The real outcome of this implemented investment will then be realized for the winners. The amount of €75, which the user chose to not invest, was paid out. While the four projects were fictional projects (because they were randomly either pro-social or for-profit projects, none of these exactly existed in reality), we invested the money according to the preferences of the participant, as good as possible. Pro-social projects were invested on Kiva and for-profit on Mintos.

Overall, we recruited 264 participants during the period of 2022.08.15 - 2022.08.26. First, we had to exclude 17 observations from the *filter* treatment, since our data showed that one participant took part at the experiment 17 times. Unfortunately, this resulted in a slightly unequal distribution of sample sizes across treatments. Based on the duration of the questionnaire, we calculated the relative speed index and excluded participants (n=13) who were above the threshold of 2 (i.e., they completed the questionnaire twice as fast as the typical one), the threshold was suggested by Leiner (2019). 234 participants remained in the final sample (female 60%, male 40%, 0% diverse; *mean* age = 27 with *SD*: 6.8).

We applied partial least squares structural equation modeling (PLS-SEM) to analyze our theoretical model. While advantages and disadvantages between this approach and the covariance-based (CB) approach are debated in current research, (e.g., Aguirre-Urreta and Marakas, 2014; Goodhue et al., 2012; Hair et al., 2011; McIntosh et al., 2014; Rönkkö and Evermann, 2013). We are bound by the nature of our model to PLS-SEM. The main reason for this is that PLS-SEM allows for non-continuous variables, while CB-SEM does not. For example, the exogenous variable for the experimental condition is a categorical variable. Furthermore, the goal of this work is to identify the existence and direction through the manipulation of anthropomorphism on user behavior. Accordingly, "if one is [...] concerned more with identifying potential relationships than the magnitude of those relationships, then regression or PLS would be appropriate" (Goodhue et al., 2012, p. 999). For the above reasons, we deem PLS-SEM to be especially suited for testing our hypotheses.

Table 3.1: Convergent and discriminant validity

Latent Construct	Cron-bach's α	CR	AVE	1	2	3	4	5	6	7
1. Perceived anthro.	0.916	0.939	0.691	0.831[a]						
2. Social presence	0.941	0.956	0.811	0.529***	0.901[a]					
3. Importance of emp.	0.749	0.842	0.571	0.321***	0.456***	0.756[a]				
4. Trusting beliefs	0.921	0.963	0.928	0.147**	0.340***	0.311***	0.963[a]			
5. Feeling observed	0.728	0.832	0.552	0.294***	0.334***	0.306***	0.082	0.743[a]		
6. Service enc. sat.	0.900	0.938	0.834	0.131**	0.369***	0.279***	0.663***	0.065	0.913[a]	
7. Enjoyment	0.940	0.962	0.894	0.319***	0.422***	0.349***	0.473***	0.185***	0.575***	0.946[a]

a The square root of the AVE is shown in the diagonal.
The lower triangle shows the correlations between the constructs: *p < 0.1, ** p < 0.05, *** p < 0.01

For each multi-latent construct, we examined the convergent and discriminant validity of the measurement instruments. The Cronbach's alphas and composite reliabilities (CR) were greater than the suggested threshold of 0.70 (Hair et al., 2016), and the values of the average variance extracted (AVE) were above the suggested minimum of 0.50 (Hair et al., 2016) (see Table 3.1). To test the discriminant validity, we assessed the factor loadings and cross-loadings (Gefen and Straub, 2005). All of the factors loaded higher on the assigned theoretical construct than on any other factor. An additional criterion for establishing discriminant validity demands that the square root of the AVE be larger than any correlation with another construct (Fornell and Larcker, 1981). This criterion was also satisfied. We concluded with the HTMT criterion, which is smaller than the threshold of 0.85 (Henseler et al., 2015). In sum, we concluded that our measures exhibited an adequate level of convergent and discriminant validity.

3.4 Results

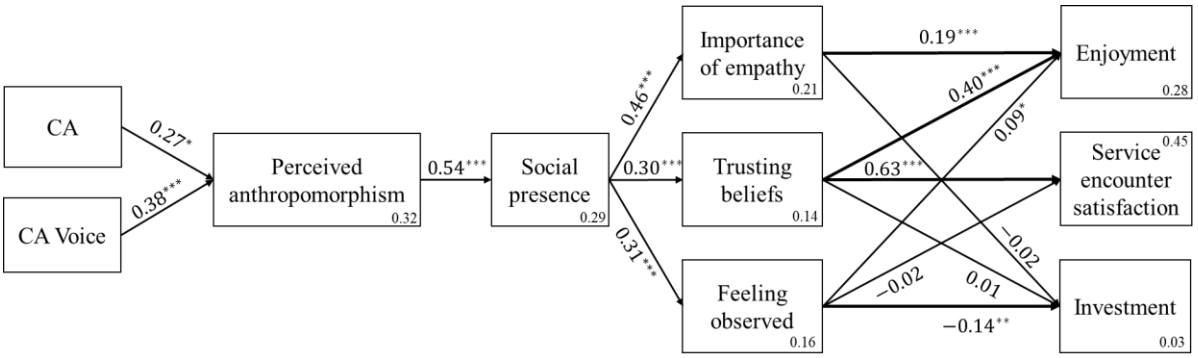
Table 3.2 summarizes the descriptive statistics, and Table 3.3 along Figure 3.5 depict the results of our statistical analyses.

Table 3.2: Descriptive statistics by treatment

Variable	Filter (N = 66)		CA (N = 80)		CA Voice (N = 88)	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*
Perceived anthropomorphism	5.15	1.51	5.62	1.19	5.84	1.26
Social presence	2.72	1.57	2.89	1.63	3.13	1.63
Importance of empathy	3.95	1.42	4.04	1.11	3.87	1.37
Trusting beliefs	4.64	1.02	5.56	0.98	5.43	1.14
Feeling observed	3.03	1.52	2.96	1.41	2.75	1.32
Enjoyment	4.28	1.38	4.39	1.46	4.39	1.55
Service encounter satisfaction	5.01	1.42	5.59	0.96	5.53	1.07
Investment	50.21	23.81	55.75	23.71	48.92	22.58

N = 234

In our model, we also controlled for our mentioned control variables: dispositional anthropomorphism, dispositional trust, social desirability bias, income and risk aversion. Our results are robust regarding excluding control variables and including the previous excluded 30 observations because of low RSIs.



Control variables: Dispositional anthropomorphism and trust, social desirability, income, risk.
Note: We used bootstrapped bias corrected confidence intervals with 5.000 samples, following the recommendation of Preacher and Hayes (2004, 2008), calculated with Smart-PLS 4.0.7.
The number in the rectangles corresponds to R^2. * p < 0.1, ** p < 0.05, *** p < 0.01

Figure 3.5: Empirical Model

Our model (see Figure 3.5 and Table 3.3) shows that the *CA Voice* treatment increases perceived anthropomorphism (H1) compared to the control treatment *Filter*, so does the CA treatment but only if tested on a 10% level. In order to also test the difference between the *CA* and the *CA Voice* treatments in perceived anthropomorphism, we conducted a one-way ANOVA combined with a Tukey HSD post-hoc test which yielded significant differences between two of three conditions (Filter & CA: $t(231) = 2.17$ $p = 0.08$; Filter & CA Voice: $t(231) = 3.24$ $p = 0.004$; CA & CA Voice: $t(231) = 1.07$ $p = 0.53$). While we do find a significant difference between *Filter* and *CA Voice*, the difference between *CA* and *CA Voice* shows no significant difference. This indicates that the voice treatment does lead to more perceived anthropomorphism, but the effect is not strong enough to make a difference to the *CA* treatment, resulting in a partial rejection of H1. This might show that the voice is not such a strong social cue as expected or the power is just too low to detect the effect.

Hypothesis H2 is supported by our results: social presence ratings are higher when participants also perceived the AS as more anthropomorphic ($\beta = 0.54$; $SE = 0.05$; $p < 0.001$). Furthermore, our model also supports H3 to H5, which means that all hypotheses regarding the effect of social presence are supported. In more detail, *social presence* does lead to higher *trusting beliefs* (H3: $\beta = 0.30$; $SE = 0.07$; $p < 0.001$), a higher *importance* of *empathy* (H4: $\beta = 0.46$; $SE = 0.05$; $p < 0.001$), and higher ratings *on feeling observed* (H5: $\beta = 0.31$; $SE = 0.06$; $p < 0.001$).

The analyses regarding the *importance of empathy* on the dependent variables show ambiguous results. While H6a is not supported ($\beta = -0.02$; $SE = 0.07$; $p = 0.78$), which means participants who consider empathy as more important did not invest a higher amount. At the same time, they seem to enjoy the process more, which supports H6b1 ($\beta = 0.19$; $SE = 0.07$; $p = 0.008$) and leads to rejection of H6b2.

Regarding the effects of *trusting beliefs*, two out of three hypotheses, were supported: Higher trusting beliefs led to more *enjoyment* (H7b: $\beta = 0.40$; $SE = 0.06$; $p < 0.001$) and a higher *service satisfaction* (H7c: $\beta = 0.63$; $SE = 0.05$; $p < 0.001$) but not to a higher *investment* (H7a: $\beta = 0.01$; $SE = 0.07$; $p = 0.93$).

Table 3.3: Summary of Hypotheses

Hypotheses	Description of Hypotheses	β (p-value)	Supported?
H1	Voice $\overset{+}{\rightarrow}$ Perceived anthropomorphism		Partly
H2	Perceived anthropomorphism $\overset{+}{\rightarrow}$ Social presence	0.54 (<0.001)	Yes
H3	Social presence $\overset{+}{\rightarrow}$ Trust	0.30 (<0.001)	Yes
H4	Social presence $\overset{+}{\rightarrow}$ Importance of empathy	0.46 (<0.001)	Yes
H5	Social presence $\overset{+}{\rightarrow}$ Feeling observed	0.31 (<0.001)	Yes
H6a	Importance of empathy $\overset{+}{\rightarrow}$ Investment	-0.02 (0.780)	No
H6b	1: Importance of empathy $\overset{+}{\rightarrow}$ Enjoyment 2: Importance of empathy $\overset{-}{\rightarrow}$ Enjoyment	0.19 (0.008)	Yes No
H7a	Trust $\overset{+}{\rightarrow}$ Investment	0.01 (0.930)	No
H7b	Trust $\overset{+}{\rightarrow}$ Enjoyment	0.40 (<0.001)	Yes
H7c	Trust $\overset{+}{\rightarrow}$ Service satisfaction	0.63 (<0.001)	Yes
H8a	Feeling Observed $\overset{+}{\rightarrow}$ Investment	-0.14 (0.047)	No*
H8b	Feeling Observed $\overset{-}{\rightarrow}$ Enjoyment	0.10 (0.090)	No
H8c	Feeling Observed $\overset{-}{\rightarrow}$ Service satisfaction	-0.01 (0.770)	No

Note: * significant in opposite direction

All our hypotheses for the effect of *feeling observed* on the outcome variables have to be rejected: First, participants which felt highly observed did *invest* less money (H8a: β = -0.14; *SE* = 0.07; p = 0.047), which contradicts our expectations. In addition, we did not find a significant effect that participants with a stronger feeling of being observed did *enjoy* the process less (H8b: β = 0.10; *SE* = 0.06; p = 0.09). Surprisingly, the sign of the coefficient is even positive and significant with a p-value below 0.1, which contradicts our expectations. For the last hypotheses, we did not find any significant differences, participants did not report a higher *service satisfaction* (H8c: β = -0.01; *SE* = 0.05; p = 0.77).

We performed several additional one-way ANOVAs to dig deeper into the relation of our experimental manipulation and the dependent variables. While we could not find effects on the invested amount (Filter & CA: $t(231) = 1.43$ $p = 0.33$; Filter & CA Voice: $t(231) = -0.34$ $p = 0.94$; CA & CA Voice: $t(231) = -1.9$ $p = 0.14$) and enjoyment (Filter & CA: $t(231) = 0.43$ $p = 0.9$; Filter & CA Voice: $t(231) = 0.45$ $p = 0.9$; CA & CA Voice: $t = 0.01$ $p = 1$), we do find that *CA* and *CA Voice* did achieve a higher service satisfaction than the *filter* treatment (Filter & CA: $t(231) = 3.35$ $p = 0.003$; Filter & CA Voice: $t(231) = 3.09$ $p = 0.006$; CA & CA Voice: $t(231) = -0.36$ $p = 0.93$).

Although we did not hypothesize a relation between our experimental manipulation and the preference for pro-social projects, our experimental design allows for such comparison. In total 130 participants chose for-profit projects and 104 chose pro-social ones. We applied a logistic regression with a dummy whether a pro-social project was chosen serving as dependent variable. We used the treatment variable as the independent variables. Revealing no effect of our treatments (CA β = 0.36 p = 0.30; CA Voice β = 0.38, p = 0.265, relative to filter). A further analysis including control variables highlights a gender

effect (β = 0.57; p = 0.045), while the odds of choosing a pro-social project as female are 1.8 times as large as the odds for a male. In an explorative analyses, two-sided t-test reveals that females do have a higher count of viewed projects (female mean = 1.71; male mean = 1.33; $t(232) = 2.97$, p = 0.003) than male and also spend more time on viewing the projects (*mean* female = 169s ; *mean* male = 123s; $t(232) = 3.02$ p = 0.003). Furthermore, the *count of viewed projects* also leads to more pro-social choices (logistic regression: β = 0.31; p = 0.032). This implies that female participants spent more time on viewing the projects and that participants who view more projects tend to choose pro-social ones.

3.5 Discussion & Future Outlook

Our results paint a mixed picture, with some hypotheses being met and some not. This raises questions that we will discuss in the following. The study found no differences for perceived anthropomorphism in between the treatments *CA* and *CA Voice* (H1). However, both treatments were significantly more anthropomorphic than the *filter* treatment, it is puzzling that the voice does not add as much to anthropomorphism as expected. A deeper analysis revealed nine outliers on the anthropomorphism scale (calculated with the whiskers confidence interval). When these outliers were excluded, significant differences between *CA* and *CA Voice* were found (CA & CA Voice: $t(231) = -2.83$ p = 0.014). Since those outliers showed no other abnormalities, we did not exclude them in our analyses. There are several possible reasons why a few participants in the *CA Voice* treatment rated the assistance system as less anthropomorphic than participants in the other treatments. First, participants in the *CA Voice* treatment resulted in a median of 3.6 minutes longer completion time compared to the *CA* treatment. Second, some participants reported that the CA's reassuring behavior after each preference was distracting. Both of these arguments could lead to an unnaturally long conversation, which could have led to a lower rating. Another explanation for the non-significant difference could be the uncanny valley effect (Mori, 1970). Findings showed that more human-like robots can lead to perceptional conflicts, which led to lower likeability of the more human-like robots (Mathur and Reichling, 2016). This possibility also exists in our voice treatment since we utilized a human voice rather than a text-to-speech generated voice, which could result in a similar perceptual conflict. Future studies should focus on designing a more natural conversation to avoid unwanted side effects (e.g., limiting reassuring behavior).

Nevertheless, the theory and other empirical studies, e.g., Cho et al. (2019) showed that the voice is a social cue and perceived anthropomorphism increases. However, this indicates that the assumption, that voice is a strong social cue transmitting many hidden information, should be scrutinized. As text-to-voice technology is becoming more and more present in daily live this might prone people to create human-to-media scripts, as suggested by Gambino et al. (2020). This is one option to explain the lower effect of the social cue. In addition, the potential adaptation of these scripts also challenges our assumption that voice is a human attribute rather than a human unique attribute, which computers can possess and thus should achieve lower humanization. If voice or other cues become a uniquely human attribute, this has important implications for both research and business, as social cues based on human

attributes would become less powerful over time. Another option might be that some participants disliked the voice itself, for which we did not control.

Our results regarding social presence influencing trust and empathy are in line with previous research (e.g., Qiu and Benbasat, 2009). Our experiment shows that social presence also leads to the feeling of being observed which is, to the best of our knowledge, new and has further implications. In addition, we did show that *importance of empathy* leads to higher enjoyment, which indicates that positive feelings like warm glow rule out potential negative feelings. Furthermore, we did find that trust has a positive influence on service satisfaction and enjoyment as hypothesized.

Although we did not find the expected effects of feeling observed on our outcome variables, we did find that the effect on investment was negative while the coefficient on enjoyment had a positive sign. One reason for a positive effect of *feeling observed*, could be that people perceive the AS as another person, which on the one hand creates the illusion of not being alone, and on the other hand might raise the idea that also other people can invest money, thus reducing the pressure to invest. Latter would even explain the negative effect on the investment level, which is also negative at least from the platform's point of view. Interestingly, this contradicts the idea that self-awareness leads to higher investments, and challenges the assumption that this can only be triggered by an AS (Sah and Peng, 2015). We considered that the type of project (pro-social or for-profit) could moderate the effect between *feeling observed* and investment, but we did not find a significant moderation (Type of project: $\beta = 0.116$ $p = 0.571$).

Lastly, our further analyses showed that participants who examined more projects chose more pro-social projects. While an additional test revealed that female in general viewed more projects, it seems to be an important part to design websites in such a way, leading their customers to viewing multiple projects if pro-social choices are preferred. Similar, future researcher should be aware of this factor as it might confound their results. In future research, one might also test CAs with male voice in order to find out whether there is a gender effect or rather an effect of fitting genders of CA and user.

3.6 Conclusion

Our experiment emphasis the complexity of creating and designing an AS which fits its context. Voice might create perceived anthropomorphism, but this did not translate on the outcome variables. As Seeger et al. (2021) showed, changing only one cue alone might be not noticeable enough to be recognized in form of the perceived anthropomorphism. Indicating that developer should not focus only on one specific cue, like the voice. While most studies do ignore the classical non-anthropomorphic design of websites, we showed that using a CA at all did increase the service satisfaction, which reveals that CAs in general have a positive effect. From a research perspective, the complex nature of a higher social presence still is of high interest and needs to be researched from new perspectives that fit the respective context, like our focus on feelings like empathy and being observed that we introduced in our experiment.

Funding

Research reported in this paper was supported by Nürnberg Institut für Marktentscheidungen e.V.

3.7 References

Aguirre-Urreta, M. I. and G. M. Marakas (2014). "Research Note—Partial Least Squares and Models with Formatively Specified Endogenous Constructs: A Cautionary Note" *Information Systems Research* 25 (4), 761–778.

Allison, T., B. Davis, J. Short and J. Webb (2015). "Crowdfunding in a Prosocial Microlending Environment: Examining the Role of Intrinsic Versus Extrinsic Cues" *Entrepreneurship Theory and Practice* 39 (1), 53–73.

Allison, T. H., A. F. McKenny and J. C. Short (2013). "The effect of entrepreneurial rhetoric on microlending investment: An examination of the warm-glow effect" *Journal of Business Venturing* 28 (6), 690–707.

Andreoni, J., J. M. Rao and H. Trachtman (2017). "Avoiding the Ask: A Field Experiment on Altruism, Empathy, and Charitable Giving" *Journal of Political Economy* 125 (3), 625–653.

Araujo, T. (2018). "Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions" *Computers in Human Behavior* 85, 183–189.

Aung, T. and D. Puts (2020). "Voice pitch: a window into the communication of social power" *Current Opinion in Psychology* 33, 154–161.

Barasch, A., E. E. Levine, J. Z. Berman and D. A. Small (2014). "Selfish or selfless? On the signal value of emotion in altruistic behavior" *Journal of personality and social psychology* 107 (3), 393–413.

Barger, P. B. and A. A. Grandey (2006). "Service with a smile and encounter satisfaction: Emotional contagion and appraisal mechanisms" *Academy of Management Journal* 49 (6), 1229–1238.

Batson, C. D. (2014). *The altruism question: Toward a social-psychological answer*. 1st Edition. New York: Psychology Press.

Batson, C. D., E. R. Thompson, G. Seuferling, H. Whitney and J. A. Strongman (1999). "Moral hypocrisy: appearing moral to oneself without being so" *Journal of personality and social psychology* 77 (3), 525.

Bentley, F., C. Luvogt, M. Silverman, R. Wirasinghe, B. White and D. Lottridge (2018). "Understanding the Long-Term Use of Smart Speaker Assistants" *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2 (3), 1–24.

Biocca, F., C. Harms and J. K. Burgoon (2003). "Toward a more robust theory and measure of social presence: Review and suggested criteria" *Presence: Teleoperators and Virtual Environments* 12 (5), 456–480.

Bruton, G. D., S. Khavul and H. Chavez (2011). “Microlending in emerging economies: Building a new line of inquiry from the ground up” *Journal of International Business Studies* 42 (5), 718–739.

Cain, D. M., J. Dana and G. E. Newman (2014). “Giving Versus Giving In” *The Academy of Management Annals* 8 (1), 505–533.

Cho, E., M. D. Molina and J. Wang (2019). “The effects of modality, device, and task differences on perceived human likeness of voice-activated virtual assistants” *Cyberpsychology, Behavior, and Social Networking* 22 (8), 515–520.

Cuff, B. M., S. J. Brown, L. Taylor and D. J. Howat (2016). “Empathy: A Review of the Concept” *Emotion Review* 8 (2), 144–153.

Cyr, Head, Larios and Pan (2009). “Exploring Human Images in Website Design: A Multi-Method Approach” *MIS Quarterly* 33 (3), 539.

Cyr, D., K. Hassanein, M. Head and A. Ivanov (2007). “The role of social presence in establishing loyalty in e-Service environments” *Interacting with Computers* 19 (1), 43–56.

Davis, M. H. (1980). *A multidimensional approach to individual differences in empathy:* American Psychological Association Washington, DC.

Davis, M. H. (1983a). “Measuring individual differences in empathy: Evidence for a multidimensional approach” *Journal of personality and social psychology* 44 (1), 113.

Davis, M. H. (1983b). “The effects of dispositional empathy on emotional reactions and helping: A multidimensional approach” *Journal of personality* 51 (2), 167–184.

Davis, M. H. (2015). “Empathy and Prosocial Behavior”. In D. A. Schroeder (ed.) *The Oxford handbook of prosocial behavior*. Oxford: Oxford Univ. Press.

DeMaio, T. J. (1984). “Social desirability and survey” *Surveying subjective phenomena* 2, 257.

Duval, S. and R. A. Wicklund (1972). “A theory of objective self awareness”.

Eisenberg, N. and P. A. Miller (1987). “The relation of empathy to prosocial and related behaviors” *Psychological Bulletin* 101 (1), 91–119.

Eisenberg, N., M. Schaller, R. A. Fabes, D. Bustamante, R. M. Mathy, R. Shell and K. Rhodes (1988). “Differentiation of personal distress and sympathy in children and adults” *Developmental Psychology* 24 (6), 766.

Ekström, M. (2012). “Do watching eyes affect charitable giving? Evidence from a field experiment” *Experimental Economics* 15 (3), 530–546.

Epley, N., A. Waytz and J. T. Cacioppo (2007). “On seeing human: A three-factor theory of anthropomorphism” *Psychological Review* 114 (4), 864–886.

Feine, J., U. Gnewuch, S. Morana and A. Maedche (2019). “A Taxonomy of Social Cues for Conversational Agents” *International Journal of Human-Computer Studies* 132, 138–161.

Fogg, B. J. (2002). “Persuasive technology” *Ubiquity* 2002 (December), 2.

Fornell, C. and D. F. Larcker (1981). “Evaluating Structural Equation Models with Unobservable Variables and Measurement Error” *Journal of Marketing Research* 18 (1), 39–50.

Freud, S. (1930). *Civilization and its discontents.* Oxford, England: Hogarth.

Gambino, A., J. Fox and R. A. Ratan (2020). “Building a stronger CASA: Extending the computers are social actors paradigm” *Human-Machine Communication* 1, 71–85.

Gefen, D. and D. Straub (2003). “Managing User Trust in B2C e-Services” *e-Service Journal* 2 (2), 7–24.

Gefen, D. and D. Straub (2005). “A Practical Guide To Factorial Validity Using PLS-Graph: Tutorial And Annotated Example” *Communications of the Association for Information systems* 16.

Gefen, D. and D. W. Straub (1997). “Gender differences in the perception and use of e-mail: An extension to the technology acceptance model” *MIS Quarterly*, 389–400.

Gefen, D. and D. W. Straub (2000). “The Relative Importance of Perceived Ease of Use in IS Adoption: A Study of E-Commerce Adoption” *Journal of the Association for Information Systems* 1 (1), 1–30.

Gefen, D. and D. W. Straub (2004). “Consumer trust in B2C e-Commerce and the importance of social presence: experiments in e-Products and e-Services” *Omega* 32 (6), 407–424.

Goodhue, Lewis and Thompson (2012). “Does PLS Have Advantages for Small Sample Size or Non-Normal Data?” *MIS Quarterly* 36 (3), 981.

Guthrie, S. (1993). *Faces in the clouds*. *A new theory of religion.* New York, Oxford: Oxford University Press.

Hair, J. F., G. T. M. Hult, C. M. Ringle and M. Sarstedt (2016). *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM).* 1st Edition. Thousand Oaks: SAGE Publications Inc.

Hair, J. F., C. M. Ringle and M. Sarstedt (2011). “PLS-SEM: Indeed a Silver Bullet” *Journal of Marketing Theory and Practice* 19 (2), 139–152.

Haley, K. J. and D. M. Fessler (2005). “Nobody's watching?” *Evolution and Human Behavior* 26 (3), 245–256.

Haslam, N. (2006). “Dehumanization: an integrative review” *Personality and Social Psychology Review* 10 (3), 252–264.

Haslam, N., P. Bain, L. Douge, M. Lee and B. Bastian (2005). “More human than you: attributing humanness to self and others” *Journal of personality and social psychology* 89 (6), 937–950.

Henseler, J., C. M. Ringle and M. Sarstedt (2015). “A new criterion for assessing discriminant validity in variance-based structural equation modeling” *Journal of the Academy of Marketing Science* 43 (1), 115–135.

Herzenstein, M., S. Sonenshein and U. M. Dholakia (2011). “Tell Me a Good Story and I May Lend you Money: The Role of Narratives in Peer-to-Peer Lending Decisions” *Journal of Marketing Research)* 48 (SPL), 138-149.

Heßler, P. O., J. Pfeiffer and S. Hafenbrädl (2022). “When Self-Humanization Leads to Algorithm Aversion” *Business & Information Systems Engineering* 64 (3), 275–292.

Holbrook, A. L., M. C. Green and J. A. Krosnick (2003). “Telephone versus face-to-face interviewing of national probability samples with long questionnaires: Comparisons of respondent satisficing and social desirability response bias” *Public Opinion Quarterly* 67 (1), 79–125.

Jones, E. and T. Pittman (1982). “Toward a general theory of strategic self-presentation” *Psychological Perspectives on the Self* 1.

Kahneman, D. and A. Tversky (2013). “Prospect Theory: An Analysis of Decision Under Risk”. In *Handbook of the Fundamentals of Financial Decision Making*, pp. 99–127: WORLD SCIENTIFIC.

Knapp, M. L. and J. A. Daly (2011). *The SAGE Handbook of Interpersonal Communication:* SAGE Publications.

Krebs, D. (1975). “Empathy and altruism” *Journal of personality and social psychology* 32 (6), 1134.

Lankton, N., D. H. McKnight and J. Tripp (2015). “Technology, Humanness, and Trust: Rethinking Trust in Technology” *Journal of the Association for Information Systems* 16 (10), 880–918.

Leiner, D. J. (2019). “Too Fast, too Straight, too Weird: Non-Reactive Indicators for Meaningless Data in Internet Surveys”. 229-248 Pages / Survey Research Methods, Vol 13 No 3 (2019).

Linek, S. B., P. Gerjets and K. Scheiter (2010). “The speaker/gender effect: does the speaker’s gender matter when presenting auditory text in multimedia messages?” *Instructional Science* 38 (5), 503–521.

Loewenstein, G. and D. A. Small (2007). “The scarecrow and the tin man: The vicissitudes of human sympathy and caring” *Review of general psychology* 11 (2), 112–126.

Luhmann, N. (2018). *Trust and power:* John Wiley & Sons.

Mathur, M. B. and D. B. Reichling (2016). “Navigating a social world with robot partners: A quantitative cartography of the Uncanny Valley” *Cognition* 146, 22–32.

Mayer, R. E., K. Sobko and P. D. Mautone (2003). “Social cues in multimedia learning: Role of speaker’s voice” *Journal of educational Psychology* 95 (2), 419.

McIntosh, C. N., J. R. Edwards and J. Antonakis (2014). “Reflections on partial least squares path modeling” *Organizational Research Methods* 17 (2), 210–251.

McKnight, D. H., V. Choudhury and C. Kacmar (2002a). “Developing and Validating Trust Measures for e-Commerce: An Integrative Typology” *Information Systems Research* 13 (3), 334–359.

McKnight, D. H., V. Choudhury and C. Kacmar (2002b). “The impact of initial consumer trust on intentions to transact with a web site: a trust building model” *The Journal of Strategic Information Systems* 11 (3-4), 297–323.

Meyerson, D., K. E. Weick and R. M. Kramer (1996). “Swift trust and temporary groups”. In R. M. Kramer and T. R. Tyler (eds.) *Trust in organizations*. *Frontiers of Theory and Research*, pp. 261–287. London: SAGE.

Moon, J.-W. and Y.-G. Kim (2001). “Extending the TAM for a World-Wide-Web context” *Information & Management* 38 (4), 217–230.

Mori, M. (1970). “The uncanny valley” *Energy* 7, 33–35.

Nass, C. and Y. Moon (2000). “Machines and mindlessness: Social responses to computers” *Journal of Social Issues* 56 (1), 81–103.

Nass, C., Y. Moon and N. Green (1997). “Are machines gender neutral? Gender-stereotypic responses to computers with voices” *Journal of applied social psychology* 27 (10), 864–876.

Paulhus, D. L. (1986). “Self-Deception and Impression Management in Test Responses”. In A. Angleitner and J. S. Wiggins (eds.) *Personality assessment via questionnaires*. *Current issues in theory and measurement*, pp. 143–165. Berlin: Springer-Verlag.

Pfeiffer, J., I. Benbasat and F. Rothlauf (2014). “Minimally restrictive decision support systems”.

Poushneh, A. (2021). “Humanizing voice assistant: The impact of voice assistant personality on consumers’ attitudes and behaviors” *Journal of Retailing and Consumer Services* 58, 102283.

Preacher, K. J. and A. F. Hayes (2004). “SPSS and SAS procedures for estimating indirect effects in simple mediation models” *Behavior research methods, instruments, & computers : a journal of the Psychonomic Society, Inc* 36 (4), 717–731.

Preacher, K. J. and A. F. Hayes (2008). “Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models” *Behavior research methods* 40 (3), 879–891.

Qiu, L. and I. Benbasat (2009). “Evaluating Anthropomorphic Product Recommendation Agents: A Social Relationship Perspective to Designing Information Systems” *Journal of Management Information Systems* 25 (4), 145–182.

Reeves, B. and C. Nass (1996). “The media equation: How people treat computers, television, and new media like real people” *Cambridge, UK* 10, 236605.

Rhim, J., M. Kwak, Y. Gong and G. Gweon (2022). “Application of humanization to survey chatbots: Change in chatbot perception, interaction experience, and survey data quality” *Computers in Human Behavior* 126, 107034.

Richter, D., M. Metzing, M. Weinhardt and J. Schupp (2013). *SOEP scales manual*. SOEP Survey Papers.

Rönkkö, M. and J. Evermann (2013). “A Critical Examination of Common Beliefs About Partial Least Squares Path Modeling” *Organizational Research Methods* 16 (3), 425–448.

Rzepka, C., B. Berger and T. Hess (2021). “Voice Assistant vs. Chatbot – Examining the Fit Between Conversational Agents’ Interaction Modalities and Information Search Tasks” *Information Systems Frontiers*, 1–18.

Sah, Y. J. and W. Peng (2015). “Effects of visual and linguistic anthropomorphic cues on social perception, self-awareness, and information disclosure in a health website” *Computers in Human Behavior* 45, 392–401.

Sargeant, A. (1999). “Charitable Giving: Towards a Model of Donor Behaviour” *Journal of Marketing Management* 15 (4), 215–238.

Scherer, K. R. (1995). “Expression of emotion in voice and music” *Journal of Voice* 9 (3), 235–248.

Schild, C., J. Stern and I. Zettler (2019). “Linking men‘s voice pitch to actual and perceived trustworthiness across domains” *Behavioral Ecology* 31, 164–175.

Schuetzler, R. M., J. S. Giboney, G. M. Grimes and J. F. Nunamaker (2018). “The influence of conversational agent embodiment and conversational relevance on socially desirable responding” *Decision Support Systems* 114, 94–102.

Schuller, B., S. Steidl, A. Batliner, A. Vinciarelli, K. Scherer, F. Ringeval, M. Chetouani, F. Weninger, F. Eyben, E. Marchi, M. Mortillaro, H. Salamin, A. Polychroniou, F. Valente and S. Kim (2013). “The INTERSPEECH 2013 computational paralinguistics challenge: Social signals, conflict, emotion, autism”. In: *14th Annual Conference of the International Speech Communication Associatio*. Ed. by INTERSPEECH 2013, pp. 148–152.

Seeger, A.-M., J. Pfeiffer and A. Heinzl (2021). “Texting with human-like conversational agents: Designing for anthropomorphism” *Journal of the Association for Information Systems* (22(4)), 931-967.

Stern, J., C. Schild, B. C. Jones, L. M. DeBruine, A. Hahn, D. A. Puts, I. Zettler, T. L. Kordsmeyer, D. Feinberg, D. Zamfir, L. Penke and R. C. Arslan (2021). “Do voices carry valid information about a speaker’s personality?” *Journal of Research in Personality* 92, 104092.

Swangnetr, M. and D. B. Kaber (2013). “Emotional State Classification in Patient–Robot Interaction Using Wavelet Analysis and Statistics-Based Feature Selection” *(EEE Transactions on Human-Machine Systems* 43 (1), 63–75.

van Kleef, G. A. (2010). “The Emerging View of Emotion as Social Information” *Social and Personality Psychology Compass* 4 (5), 331–343.

Verhagen, T., J. van Nes, F. Feldberg and W. van Dolen (2014). “Virtual Customer Service Agents: Using Social Presence and Personalization to Shape Online Service Encounters” *Journal of Computer-Mediated Communication* 19 (3), 529–545.

Waytz, A., J. Cacioppo and N. Epley (2010a). “Who Sees Human? The Stability and Importance of Individual Differences in Anthropomorphism” *Perspectives on Psychological Science* 5 (3), 219–232.

Waytz, A., J. Heafner and N. Epley (2014). “The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle” *Journal of Experimental Social Psychology* 52, 113–117.

Waytz, A., C. K. Morewedge, N. Epley, G. Monteleone, J.-H. Gao and J. T. Cacioppo (2010b). “Making sense by making sentient: effectance motivation increases anthropomorphism” *Journal of Personality and Social Psychology* 99 (3), 410–435.

Weinstein, N. and R. M. Ryan (2010). "When helping helps: autonomous motivation for prosocial behavior and its influence on well-being for the helper and recipient" *Journal of personality and social psychology* 98 (2), 222–244.

Winkler, N., M. Kroh and M. Spiess (2006). *Entwicklung einer deutschen Kurzskala zur zweidimensionalen Messung von sozialer Erwünschtheit.* DIW Discussion Papers.

Yoo, Y. and M. Alavi (2001). "Media and Group Cohesion: Relative Influences on Social Presence, Task Participation, and Group Consensus" *MIS Quarterly* 25 (3), 371–390.

Zwebner, Y. and R. Y. Schrift (2020). "On My Own: The Aversion to Being Observed during the Preference-Construction Stage" *Journal of Consumer Research.*

3.8 Supplemental Material

Table 3.4: Overview over scales rom questionnaire

Scale	**Source/Items**
Perceived Anthropomorphism	See Waytz et al. (2010b).
Social Presence	See Gefen and Straub (1997; 2003).
Importance of Empathy	See Davis (1980, 1983a) and Heßler et al. (2022).
Trusting beliefs	See Lankton et al. (2015).
Feeling observed	When making the investment decision, I had... ... the feeling of being watched by the assistance system. ... thinking about how others would have decided. ... thought about how others would judge my decision. ... the feeling that the assistance system was following my decision.
Enjoyment	See Moon and Kim (2001).
Service encounter Satisfaction	See Barger and Grandey (2006).
Social desirability bias	See Winkler et al. (2006) and (Paulhus, 1986).
Dispositional trust	See McKnight et al. (2002a).
Dispositional anthropomorphism	See Waytz et al. (2010a).
Risk aversion	See Richter et al. (2013).

Note: If needed, scales were rephrased to fit our context and translated to German for the experiment.

4 Paper C: The Voice Effect

Rethinking Gender Stereotypes in Conversational Agent Design

Pascal Oliver Heßler • Jella Pfeiffer • Matthias Unfried

Abstract

This study explores how the gender and voice of conversational agents (CAs) affect user behavior, particularly in decisions involving pro-social versus for-profit preferences. While prior research and an earlier study (Heßler et al., 2023) suggested that gender stereotypes might influence outcomes, our findings reveal a different pattern: the presence of a voice, rather than its gender, significantly impacts men's choices. This voice effect challenges prevailing assumptions about gendered CAs and underscores the importance of designing systems that avoid reinforcing stereotypes while ensuring usability and engagement.

4.1 Introduction

The rapid development of CAs and advances in voice synthesis technology have revolutionized human-computer interaction. From their initial implementation with limited capabilities, CAs now integrate sophisticated features, including the ability to converse in natural language. Through advancements in synthesizing voices, many CAs can now speak. In parallel, different representations of CAs were developed, from robots to animals to humans assigned a specific gender, often reflected through their voice, name, or appearance.

Many CAs, such as Amazon's Alexa, Apple's Siri, or Google's Google Assistant, are designed with gendered attributes, typically defaulting to female voices and names (Abercrombie et al., 2021). While most companies nowadays provide the possibility to change the voice to a male voice, the default setting often remains female. Apple has introduced greater flexibility by not setting a default voice, instead letting users choose between male and female voices (Fournier-Tombs, 2021). Regardless of the voice, the names of the CAs often remain female, such as Alexa, Cortana, and Siri. This trend has sparked widespread criticism, with concerns that such practices reinforce traditional gender stereotypes and propagate harmful societal norms. For instance, UNESCO has highlighted that the submissive and accommodating personas frequently attributed to female CAs could reinforce biases about women as subservient or compliant (West et al., 2019).

Besides the mere criticism, research shows that such stereotypes significantly shape humans' perceptions and behavior. Stereotyping is defined as generalized beliefs about specific groups (Heilman, 2012) and influences decision-making processes, often unconsciously (Bargh et al., 1996; Dijksterhuis & Bargh, 2001). Gender stereotypes, in particular, shape attitudes toward men and women, and their application to CAs raises critical ethical and design concerns. Research has shown that assigning a specific gender to a CA can evoke stereotypical reactions from users, influencing both their perceptions

and interactions with the system (Ahn et al., 2022; Liao & Huang, 2024). These reactions may yield positive effects, such as increased trust, but can also have adverse effects, including the reinforcement of harmful stereotypes.

While studies investigating the use of gender stereotypes in CAs exist, most do not explore how the user's gender, in combination with the CA's gender, might affect user behavior (e.g., Ahn et al., 2022; Liao & Huang, 2024). Studies often concentrate on the direct effect of male or female voices. For example, Schild et al. (2019) found that a male voice engenders greater trust, while Linek et al. (2010) observed that female voices improved learning outcomes in educational contexts.

Researching only CA's gender effects seems logical at first since gender stereotypes are not dependent on the gender of the person perceiving them, as these biases can influence all of us (Mahmood & Huang, 2024). Nevertheless, some studies suggest that matching gender between users and CAs may impact user perceptions. For instance, research has shown that participants tend to trust voices of the same gender more (S. K. Lee et al., 2021). However, the interaction between the user's gender and the CA's gender is still underexplored, particularly regarding how these dynamics affect decision-making and preferences. This paper examines how the interaction between participants' gender and CA's gender influences their choices regarding pro-social or for-profit options. We selected this context because it allows us to investigate whether gender stereotypes shape participants' preferences. Heßler et al. (2022) demonstrated that the motivations behind pro-social and for-profit decisions differ, underscoring the relevance of tailoring CAs to specific contexts. Furthermore, these two decision types serve as natural counterparts, providing a clear framework for examining the role of gender stereotypes.

We acknowledge that gender is not a binary construct, and many individuals do not conform to traditional male or female roles. However, the stereotypes associated with men and women dominate societal perceptions and form the basis of how gendered CAs are typically designed and evaluated. In this context, we focus on binary gender stereotypes, as they remain central to the discourse surrounding CAs.

Based on our previous study (Heßler et al., 2023), we conducted an exploratory analysis of user responses to CAs with female voices. Our findings suggested that male users responded more strongly to female voices, exhibiting a heightened preference for for-profit projects. These unexpected results raised questions about the robustness of this effect and whether it could be replicated. While our initial study focused on the CA's social cues, like voice, we hypothesized that gender stereotypes might drive the observed effects.

To investigate these findings further, we designed a follow-up study incorporating male and female CAs to explore the interaction between user gender and CA gender with and without voice in greater depth. The gender of participants adds further complexity. Participant gender may influence how users respond to CAs, yet the interplay between user gender and CA gender remains underexplored in research. Due

to sample size and statistical power challenges, this study partly excludes participants outside the male-female binary, which we recognize as a limitation. We also include an alternative approach, where participants self-reported how they identify with masculine and feminine characteristics (BSRI-R scale); Also, we include an alternative approach; future research should aim to address these gaps, incorporating a more inclusive approach to gender dynamics in human-agent interactions.

This study examines how the interaction between the user's gender and the CA's gender influences behavior, specifically in contexts involving pro-social versus for-profit decisions. Our findings aim to contribute to a nuanced understanding of the role of gender in designing ethical and effective CAs while acknowledging the complexities and limitations inherent in such research.

Our results reveal no interaction effect between the user's gender and the CA's gender. However, we found that men are more influenced by the presence of a voice, regardless of whether it is female or male voice.

This paper is structured as follows: First, we analyze the initial study's results, reevaluating its findings and focusing on the distinction between gender and voice effects. Next, we present our hypotheses, building on these insights to frame the expectations for the new experiment. Following this, we detail the new experiment's results, highlighting the impact of voice presence and gender on user behavior. Finally, we discuss the broader implications of our findings and provide an ethical contextualization of the role of gender in CA design.

4.2 Explorative results of Heßler et al. (2023)

The following briefly summarizes the previous study from Heßler et al. (2023), highlighting our explorative findings regarding the gender effect. The study consists of five parts, and a more detailed description can be found in the original paper. Participants were given identical information on the experiment and microlending types (pro-social and for-profit). The CA first collected their preference for pro-social or for-profit projects, followed by more detailed preferences for microlending attributes (e.g., interest type, credit score, financing status, duration, and region). They then reviewed four projects: the first two aligned with their pro-social/for-profit preference, while the last two did not. Participants could invest up to €75 in one project. The session concluded with a questionnaire.

The study included three treatments, two CAs: female CA and the female CA with a voice. In addition, it had one control treatment without a CA were people could set their preferences directly. We designed the study so that participants' preference for pro-social of for-profit projects was measured twice: through self-reported preferences and through behavioral choices during project investment. In the following analysis, we will focus on both measures. However, we will only analyze the CA treatments, excluding the control treatment since no gender was assigned to it. In total, we collected 168 observations.

At first, a t-test reveals that women scored 0.61 points lower on the self-reported scale, indicating a stronger preference toward pro-social options (mean women = 3.447; mean men = 4.054; $t(232) = -3.39$, $diff = -0.607$, $p < 0.001$). In the next step, we analyzed whether there was a treatment effect that affected male and female participants differently. For this, we utilized ANOVA to compare the treatments and the gender, as shown in Figure 4.1. The analysis identified a significant difference in the voice treatment between women and men, with men showing a stronger preference for for-profit projects. These results indicate that men were influenced not directly by the CA's gender but by the female CA voice.

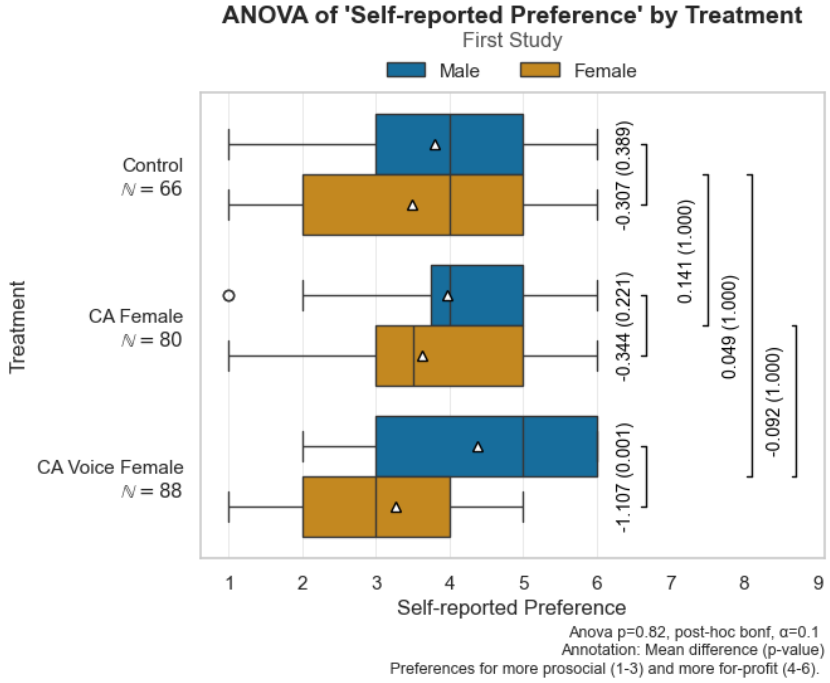


Figure 4.1: Box-Plot Self-reported Preference by treatment and gender

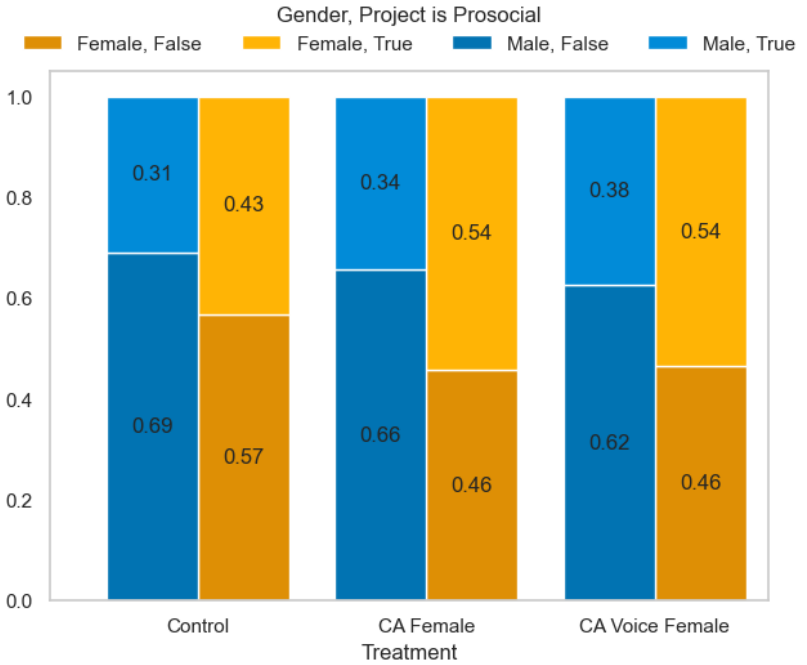


Figure 4.2: Relative count of chosen project type by treatment and gender for study 1

When looking at the actual decisions, we did not find the same effect (Figure 4.2). While women continued to choose more pro-social projects, their choices were not influenced by the CA. In contrast, men selected more for-profit projects when interacting with the CA, particularly when a voice was present. Given the binary nature of the data, we conducted a logit regression that included the treatment, gender of the participant, and their interaction. We did not find any significant effects. It remains unclear whether our CAs influence participants' final decisions.

In summary, we observed significant differences in the preferences but no significant differences in final decisions. This leaves some important questions unanswered. The biggest drawback of our results is the exclusive use of a female CA, making it unclear whether the observed effects are driven by the voice itself or the interaction between the CA's voice and gender.

4.3 Hypotheses Development

Building on the findings of our previous study, we aim to explore why the observed effects may occur, forming the basis of our hypotheses. The similarity-attraction theory posits that individuals are naturally inclined to trust and feel more attracted to those who share similar characteristics (Byrne, 1997; Ertug et al., 2022). This theory is backed by many researchers and led to similar and new frameworks, like homophily (Ertug et al., 2022), reducing interaction uncertainty (S. K. Lee et al., 2021; Mahmood & Huang, 2024). These frameworks suggest that shared traits, such as gender alignment between users and CAs, may enhance user trust and engagement, providing a theoretical grounding for our hypotheses.

Gender, as a salient similarity cue, has been shown to influence perceptions of trustworthiness in social interactions. For instance, individuals perceive synthesized speeches from their own gender as more attractive and trustworthy (E. J. Lee et al., 2000). The authors argue that matching gender might trigger an identification process, leading to such effects.

While some contradictory findings exist, such as evidence that trust can increase for opposite-gender interactions (Siegel et al., 2009), the dominant body of literature supports the notion that similarity strengthens trust. Even though our previous study suggests otherwise, we follow the notion of the literature. Gender similarity, in particular, has been identified as a salient factor influencing interpersonal dynamics and decision-making processes. Therefore, we hypothesize:

H1: Matching the user's gender with the CA's gender increases trust towards the CA.

While the CA's gender influences how users interact with CAs, these effects are not uniform across users' genders (Christov-Moore et al., 2014; Kraus et al., 2018; Nass et al., 1997; Sargeant, 1999). Research indicates that women tend to be more skeptical of CAs, regardless of the CA's gender (S. K. Lee et al., 2021; Ruijten et al., 2015). This skepticism may stem from higher expectations (Ma et al., 2024) and a more critical evaluation of the perceived usefulness of recommendations (Doong & Wang,

2011). As a result, the gender of the CA is less likely to influence women's preferences or decisions strongly.

In contrast, male participants seem to be more affected by the gender of the CA, particularly when interacting with female agents. This phenomenon aligns with findings from Siegel et al. (2009), where men exhibited greater trust in female robots than their male counterparts. The explanation for this difference could be linked to gendered perceptions of warmth and social desirability, where users perceive female agents as more supportive and agreeable (Kraus et al., 2018; Mahmood & Huang, 2024). This aligns with broader social dynamics in which men may be more susceptible to gendered cues emphasizing emotional warmth and care, both of which are qualities stereotypically associated with women.

Overall, these dynamics suggest that male participants are more likely to be influenced by the gender of the CA than female participants, leading to the following hypothesis:

H2: Men are more influenced by the gender of the CA than women.

Since trust decreases uncertainty, this should also positively affect the investment amount. As investments inherently involve risk, trust in the CA should also spill over to the final project decision, whether it is a for-profit or pro-social project.

Furthermore, while men appear to donate less frequently, some studies suggest they contribute larger amounts than women (De Wit & Bekkers, 2016; Braus, 1994, quoted from Sargeant, 1999). Conversely, other research indicates that women donate more to charitable causes (De Wit & Bekkers, 2016; Mesch et al., 2011). Given these conflicting findings, the relationship between gender and donation behavior remains unclear. Therefore, we propose only a hypothesis that trust increases investment and include a control variable for participants' gender to account for potential differences.

H3: More trust in the CA leads to a higher investment amount.

While the previous hypothesis centers on the investment amount, we now focus on preferences and how the CA's gender might influence them. Gender stereotypes are pivotal in shaping perceptions of CAs (Angeli & Brahnam, 2006). Research consistently shows that women are perceived as more empathetic and warm, whereas men are associated with traits like competence and trustworthiness (De Wit & Bekkers, 2016; Kraus et al., 2018). Interestingly, these stereotypes often lead to a perceived lack of complementary traits, with women being viewed as less competent and men as less warm (Angeli & Brahnam, 2006).

Building on this, the interaction between in-group and out-group dynamics may further explain how users evaluate CAs. Social identity theory posits that individuals tend to favor members of their in-group while evaluating out-group members through the lens of stereotype-driven expectations (Tajfel & Turner, 2004). From this perspective, men interacting with a female CA might view it or rather her as

part of an out-group due to the gender difference. Research suggests that out-group members may evoke envy or admiration for traits perceived as stereotypically stronger in that group (Cuddy et al., 2008). In this case, men may assign greater importance to empathy when interacting with a female CA, recognizing it as a trait that women value.

Conversely, for women, a female CA represents an in-group member. Women may already expect empathy as a shared characteristic within their group and thus may not place additional emphasis on this trait when interacting with a female CA. This distinction in how in-group and out-group dynamics influence the perceived importance of empathy aligns with previous findings on stereotype application, where the salience of a trait is more pronounced when it is perceived as complementary to one's own characteristics (Cuddy et al., 2008; Heilman, 2012). These considerations lead us to the following hypotheses:

H4a: Men interacting with a female CA place greater importance on empathy than when interacting with a male CA.

H4b: Women do not place greater importance on empathy when interacting with a male or female CA.

While H4 highlights how empathy and in-group/out-group dynamics may influence users' evaluations of a CA, these dynamics extend to broader decision-making contexts. Gender stereotypes influence how empathy is perceived and, in turn, affect preferences for pro-social versus for-profit decisions. Empathy, often stereotypically attributed to women, is a critical driver of pro-social behavior, such as charitable donations (Berman et al., 2018; Eisenberg & Miller, 1987), including the willingness to donate to charities. Research shows that more empathetic individuals, such as stereotypical women, are more likely to engage in pro-social acts like charitable causes (Coke et al., 1978; Eisenberg & Miller, 1987; Verhaert & van den Poel, 2011). Since studies have demonstrated that humans interact with CAs (or computers in general) as if they were humans (Nass et al., 1994; Nass & Moon, 2000), users are likely to attribute human-like qualities, including stereotypical traits, to gendered CAs. The warmth and perceived empathy of a female CA may serve as persuasive social cues, as warmth has been shown to increase liking and trust in human interactions (Fiske et al., 2007). Similarly, these cues could lead users to align with pro-social preferences when engaging with a female CA, particularly in emotionally or socially charged contexts. Consequently, a female CA's perceived empathy and warmth may elicit stronger pro-social preferences from users, particularly in contexts involving charities.

The gender-occupational role the CA takes must also be considered, as research has already highlighted that we prefer a matching role (Tay et al., 2013, 2014). Assistant roles, such as those fulfilled by CAs, are frequently occupied by female CAs. Feine et al. (2020) showed that over 70% of the names and avatars used for CAs are classified as female. Since users are accustomed to female assistants in many assistant-like systems, they may perceive female CAs as the "default" for these roles, leading to more

favorable responses in contexts where empathy and warmth are desirable (Loideain & Adams, 2020; Tolmeijer et al., 2021). Therefore, female CAs are likely to meet user expectations more effectively, which could amplify their positive effects on pro-social preferences. In contrast, male CAs, while perceived as competent and trustworthy, may lack the empathetic qualities that align closely with pro-social decision-making, leading to a stronger for-profit preference. Building on this, we propose the following hypothesis:

H5: Female CA increases preference for pro-social projects relative to male CA.

The anthropomorphization of CAs enhances their perceived human likeness, significantly influencing user interactions. We define anthropomorphism as the attribution of human characteristics to non-human entities (Epley et al., 2007; Guthrie, 1993). While assigning a gender to a CA is already a form of anthropomorphization, we aim to further enhance it by incorporating voice as well, as it serves as a powerful social cue that conveys information about the agent and facilitates the attribution of human-like traits (Nass et al., 1997; Nass & Moon, 2000). For instance, a voice can reveal details about its "owner," such as their origin, emotional state, and other personal characteristics (Aung & Puts, 2020; Scherer, 1995; Stern et al., 2021). Adding more social cues increases anthropomorphization, which in turn fosters greater trust (Cyr et al., 2009; Gefen et al., 2003; Qiu & Benbasat, 2009).

Since voice cues also encompass gender and associated cues (e.g., emotions), we argue that they should enhance the salience of gender stereotypes. Users attribute stereotypical qualities—such as warmth for female voices—based on vocal characteristics (Mitchell et al., 2011). Similarly, Nass et al. (1997) already demonstrated that people perceive male and female voices differently, aligning with gender-stereotypical expectations. These attributions reinforce broader stereotypes and suggest that voice amplifies the effects of a CA's perceived gender on user preferences. For example, users may perceive a female CA with a recorded female voice as even warmer and more empathic, thereby reinforcing pro-social preferences.

Nonetheless, it must be noted that there are some inconsistencies in the evidence. Kraus et al. (2018) found that synthesized voice did not significantly influence user interactions in stereotypical domains, raising questions about the impact of voice alone. This discrepancy may be attributed to differences between synthesized and pre-recorded voices. Synthesized voices may lack the richness and nuance of real human voices, limiting their ability to evoke strongly gendered expectations. To address this limitation, the current study uses pre-recorded voices, which more closely mimic natural human speech and are expected to enhance anthropomorphization effects.

Voice serves as a moderator by increasing the realism of the CA and thereby strengthening the influence of the CA's gender on project preferences. Specifically, we hypothesize that the presence of voice will enhance the stereotypical expectations associated with gendered CAs. This means a female CA with a voice should elicit stronger pro-social preferences than a female CA without a voice. Similarly, while

we do not predict a specific directional effect for male CAs, we expect voice to amplify whichever traits the user associates with the male CA. This leads us to the following hypothesis:

H6: Anthropomorphization moderates the effect of the CA's gender on project preference.

4.4 Method

4.4.1 Experimental Design and Procedure

The design is mostly the same as in the previous paper (Heßler et al., 2023); please refer to the paper for a more detailed description. In the following, we only describe the changes we applied.

The most important change is adding two additional treatments for the male version: a male chatbot without and with a prerecorded voice. In conclusion, we had five treatments, including one control treatment.

a) Female version of the CA b) Male version of the CA

Figure 4.3: First message of female and male CA version

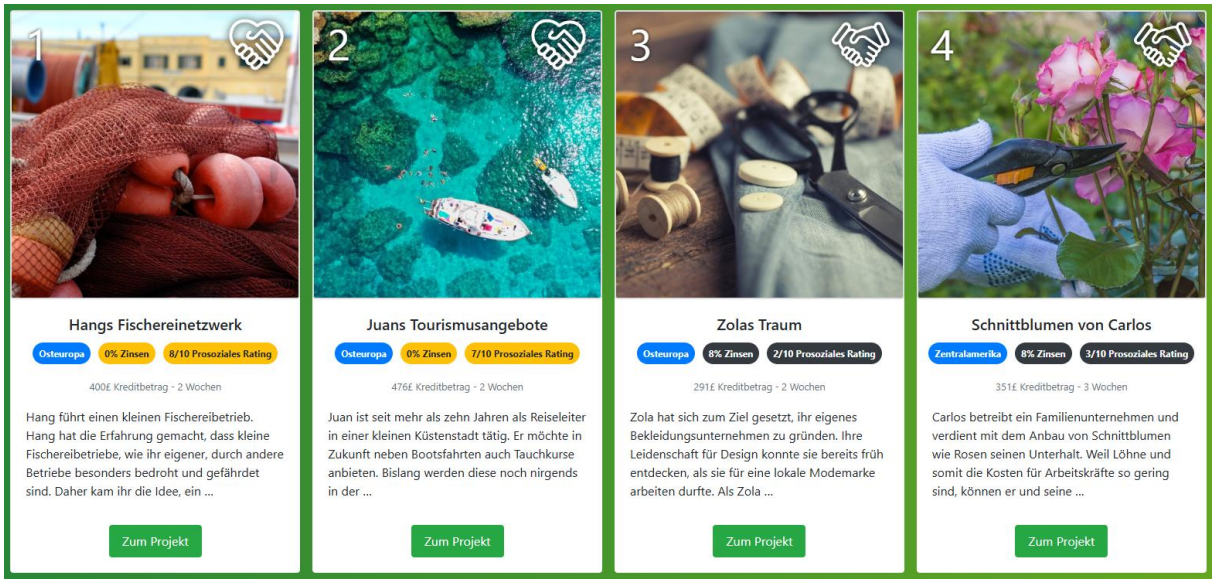


Figure 4.4: Example of the four different projects

Besides this major change, we optimized the overall experience. For example, we now show the four different projects side-by-side, rather than stacked vertically. This adjustment ensures that all projects

are visible simultaneously, eliminating the need for participants to scroll to view additional options. We also added more visual cues to indicate the type of project (pro-social or for-profit) to ensure that people were aware when certain projects did not align with their stated preferences. Figure 4.4 illustrates an example for a participant with a pro-social preference: the first two projects are framed as pro-social, followed by the for-profit options. Finally, we reduced the investment amount from €75 to £65.

4.4.2 Operationalization of the Variables

While we mainly collected the same variables as in the previous study, we also introduced some new ones. The following items were retained from the first study: perceived anthropomorphization (Waytz, Morewedge, et al., 2010), social presence (Gefen et al., 2003; Gefen & Straub, 1997; Heßler et al., 2022), the importance of empathy (Heßler et al., 2022), trusting beliefs (human version) (Lankton et al., 2015), feeling observed (Heßler et al., 2023), enjoyment (Moon & Kim, 2001), service encounter satisfaction (Barger & Grandey, 2006), dispositional trust (McKnight et al., 2002), dispositional anthropomorphism (Waytz, Cacioppo, et al., 2010), and risk aversion (Richter et al., 2013). In addition, we included a question about the perceived pleasantness of the voice in the voice treatments, as swell as the Bem Sex-Role Inventory (BSRI-R) scale (Troche & Rammsayer, 2011), a single question about the perceived impact of the investment, and a question about sexual orientation in the demographic section.

4.4.3 Sample and Model Validation

Ethical approval[8] for the experiment was obtained from the GfeW; the study was also preregistered[9]. As the preregistration included additional hypotheses beyond the scope of this paper's focus on gender stereotypes, we have included the results of these additional analyses in Appendix B for transparency and completeness.

In the first study, we collected a student sample; for this study, we used Prolific's crowd-working platform. As in the previous study, participants were tested for their ability to hear audio. With the addition of two treatments, we aimed for a sample size of 500. Participants received an upfront payment of £4.5 and had an 8% chance to win a portion of a £65 budget. We screened for only fluent German speaker, living in Germany, Austria or Switzerland.

461 participants were recruited between April 4 and April 26, 2024. Two participants reporting technical difficulties were excluded (n = 459). Since gender information was required, those without it were excluded (n = 456). Additionally, we excluded seven participants who identified as non-binary due to insufficient observations and the absence of specific hypotheses for this group. The final dataset comprised 449 participants (48% female, 52% male; mean age = 28.7, SD = 8.1).

We tested each multi-latent construct's convergent and discriminant validity. For convergent validity, we analyzed Cronbach's alpha, the composite reliability, and the average variance extracted (Hair et al.,

[8] https://gfew.de/ethik/NFn4JrEV
[9] https://aspredicted.org/FXW_HQT

2016). To analyze discriminant validity, we utilized the Fornell-Lacker criterion (Fornell & Larcker, 1981) and the HTMT criterion (Henseler et al., 2015). All constructs fulfilled the necessary conditions. Appendix A includes an overview of the criteria.

4.5 Results

Table 4.1 presents a descriptive overview of the main variables by treatment. First, we performed two manipulation checks, starting with perceived anthropomorphism. The table shows that the mean values are mainly around two, signaling a low level of anthropomorphization. An ANOVA revealed no significant differences in anthropomorphization between the CAs. However, when comparing the CAs to the control treatment, most differences were significant at the 5% level. Exceptions included the CA Male treatment, which only differed at the 10% level (*diff* = 0.392 *p* = 0.074), and CA Voice Female treatment, which showed no significant difference. The second manipulation check used the social presence scale depicted in Figure 5. ANOVA results showed that all CAs exhibited significantly higher social presence than the control treatment at the 1% level. Among the CAs, there were no significant differences except between the CA Voice Male and the CA Voice Female treatments (*diff* = -0.701 *p* = 0.033). We can conclude that the CAs generally enhance perceived anthropomorphism and social presence. Interestingly, however, the presence of a voice does not appear to significantly increase either perceived anthropomorphism or social presence. In addition, the CA Voice Male treatment had the lowest social presence mean among the CAs at 2.98, which was also lower than both CAs without a voice.

Table 4.1: Descriptive statistics by treatment

		Female		**Male**	
Variable (N = 449)	**Control** (N = 88)	**CA** (N = 90)	**CA Voice** (N = 91)	**CA** (N = 90)	**CA Voice** (N = 90)
Perceived Anthro.	1.57 (0.86)	2.09 (1.10)	1.99 (1.26)	1.96 (1.06)	2.11 (1.14)
Social Presence	1.78 (1.25)	3.32 (1.48)	3.68 (1.70)	3.13 (1.56)	2.98 (1.45)
Trusting Beliefs	4.39 (0.97)	5.56 (0.88)	5.54 (1.10)	5.29 (0.90)	5.51 (0.99)
Investment	44.31 (18.43)	45.81 (16.24)	44.89 (18.24)	42.40 (18.23)	43.72 (18.51)
Enjoyment	4.27 (1.30)	5.17 (1.16)	4.99 (1.55)	4.87 (1.32)	4.47 (1.56)
Service Satisfaction	5.14 (1.06)	5.94 (0.85)	5.70 (1.35)	5.87 (0.96)	5.61 (1.22)
Self-reported Pref.	3.82 (1.30)	3.99 (1.46)	3.96 (1.48)	3.79 (1.40)	4.04 (1.45)
Project is Pro-social	50% (0.50)	41% (0.49)	40% (0.49)	47% (0.50)	36% (0.48)

Annotation: mean (std)

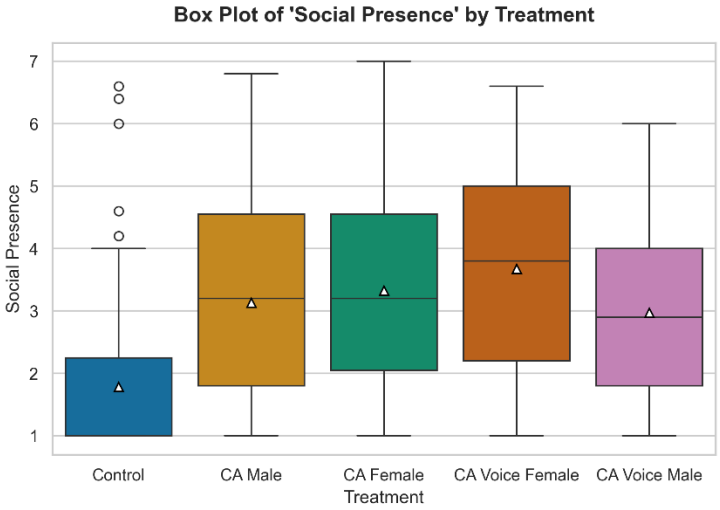
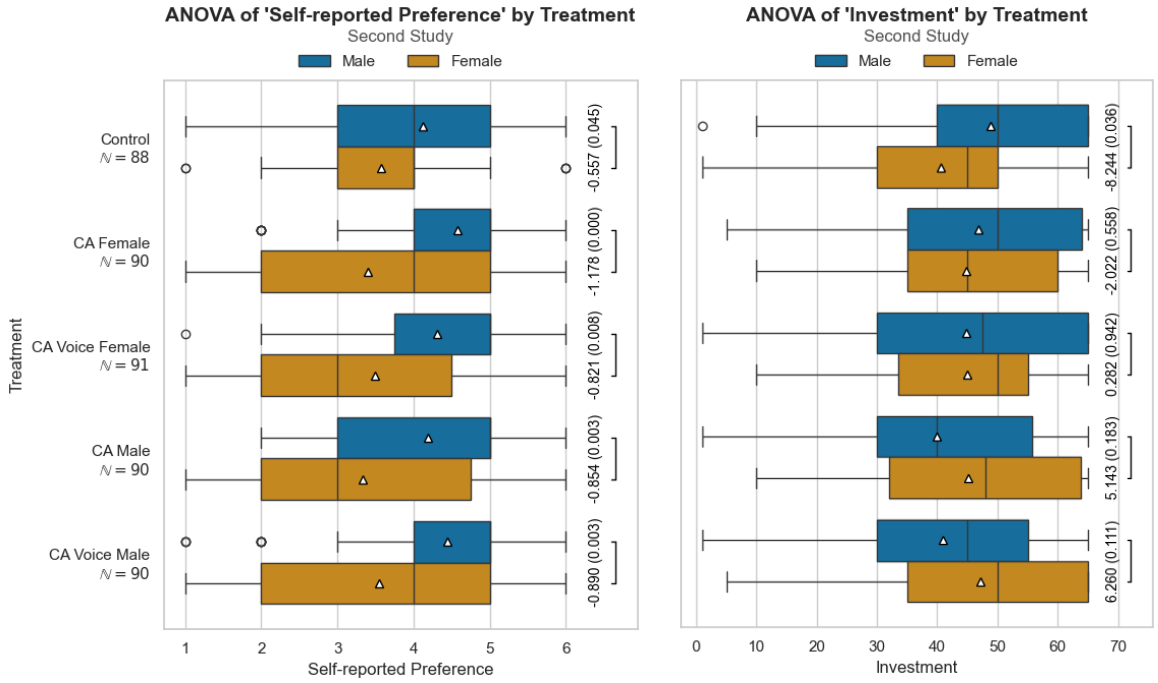


Figure 4.5: Social Presence by Treatment

Our hypotheses focus on three dependent variables: self-reported preference, whether the project is pro-social (representing behavioral preference), and the investment amount. Figure 4.6 and Figure 4.7 show a graphical overview of theses variables by treatment and participant gender. Especially Figure 4.6 a) indicates that we did not replicate the results from the first study since male participants generally showed a preference for the for-profit projects, regardless of whether they heard a voice.



a) Self-reported Preference by Treatment and Participant Gender

b) Investment by Treatment and Participant Gender

Figure 4.6: Gender Differences

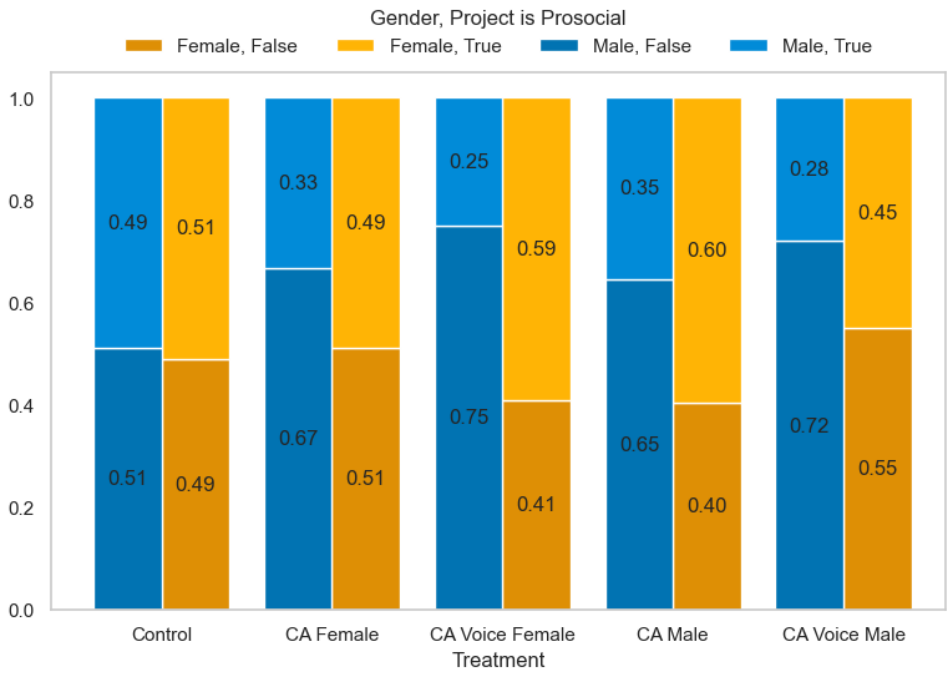


Figure 4.7: Relative count of chosen project type by treatment and gender

To test our hypotheses H1-H3, we utilized a mediation model using SmartPLS 4.1.0.9 (Ringle et al., 2024). We calculated the model depicted in Figure 4.8 and added the direct effect of gender match on investment to control for a direct effect.[10] Since we used the gender match variable, we could not include the control treatment, which had no gender, leading to a sample size of 381. The results do not support any of the hypothesized effects. The gender match does not increase trusting beliefs, failing to support H1. Male participants did not moderate the effect of gender match on trusting belief or the investment amount, leaving H2 unsupported. Finally, increased trust did not lead to higher investment amounts, failing to support H3. Interestingly, the relationship between trust and investment was even negative.

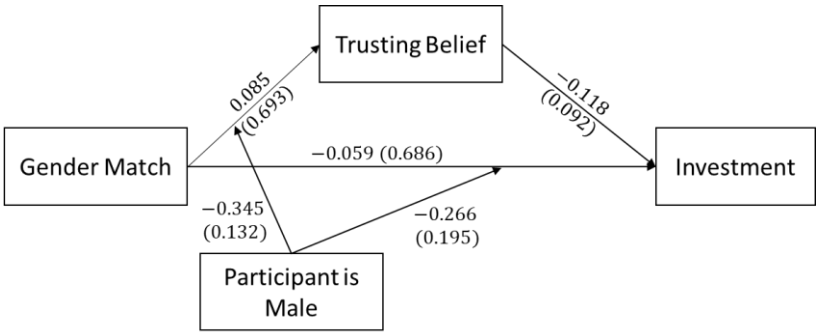


Figure 4.8: Empirical model for H1-H3

To evaluate hypotheses H4a and H4b, we conducted a simple t-test. First, we tested whether the mean importance of empathy is higher for men when interacting with a female CA compared to a male CA. The results revealed no significant difference (one-sided: $t(193) = 0.684$, $diff = 0.117$, $p = 0.2475$).

[10] We used bootstrapped bias-corrected confidence intervals with 5.000 samples.

Similarly, we ran the same test for female participants, revealing no significant differences either (two-sided: $t(193)=1.160$, $diff = 0.194$, $p = 0.124$). In summary, only H4b can be supported, indicating that the CA's gender does not affect the importance of empathy for either men or women.

H5 and H6 focused on the preference for either pro-social or for-profit projects. For the self-reported preference, we used an OLS regression. Since the final project decision is a binary variable, we used a logit regression to analyze it. The results are presented in Table 4.2. As in previous analyses, the control treatment was excluded. In both cases, an increase in the dependent variable suggests a stronger preference for pro-social projects. In the OLS regression, the coefficients can be interpreted as absoulte values, whereas in the logit regression, they are interpreted in percentage points. The results do not support H5 or H6. The only significant effect observed was through the control variable of the participant's gender. In both regressions, women decreased the dependent variable, indicating a reduced preference for for-profit projects or, conversely, an increased preference for pro-social projects.

Table 4.2: Results Preferences H5 and H6

Dependent variable	**Self-Reported Preference** β (std)	**Decision for pro-social project** β (std)
Constant	-4.561*** (0.226)	-0.932 ** (0.348)
CA is Female		
Male CA	Base	Base
Female CA	0.120 (0.298)	0.212 (0.640)
Anthropomorphism	0.108 (0.093)	0.054 (0.141)
Female CA x Anthropomorphism	-0.900 (0.128)	-0.125 (0.195)
Gender Participant		
Man	Base	Base
Women	0.934*** (0.226)	0.972*** (0.222)

Since our manipulation checks already revealed that anthropomorphism is relatively low and does not differ strongly between the treatments, we calculated the same models, including social presence, directly with a binary voice/no-voice variable. In addition, we also ran the original model with the control treatment. All results stayed robust. Finally, we calculated a model using the BSRI-R scale instead of the gender variable, including the 10 previously excluded observations from nonbinary participants. While the BSRI-R variable was significant, it explained variance in the model than the gender variable (Adjusted R^2: 0.01 vs. 0.10 of the OLS regression).

4.6 Discussion

4.6.1 Explorative Results and Discussion

First, we must acknowledge that our experiment was unable to verify most of our hypotheses, except for H4b, which postulated no effect. The only significant finding was a general gender bias among participants, independent of the treatments. Figure 6a and Figure 7 clearly illustrate this effect: men tend to prefer for-profit projects, while women, with a mean of 3.5 on the 6-Likert scale, fall between for-profit and pro-social preferences. This aligns with common stereotypes that women gravitate more toward pro-social topics than men. Interestingly, Figure 4.7 shows that the control treatment results in an almost equal split between pro-social and for-profit preferences for both men and women. However, when the CA is introduced, differences between genders become apparent, with the presence of a voice further amplifying this effect. The question remains whether these visual differences are statistically significant.

We conducted an additional logit model analysis based on Figure 4.7, incorporating treatment, participant gender, and their interaction. The results, presented in Table 4.3, reveal notable findings. Specifically, men appear to react significantly to the presence of a voice, as both the CA Voice Female and CA Voice Male treatments are associated with a higher probability of choosing a for-profit project. This suggests that the observed effect is not driven by gender matching or higher anthropomorphism but rather by the presence of a voice itself, regardless of its gender.

Although Figure 4.7 hints that this effect might extend to the no-voice treatment, the magnitude of this effect is likely too small to be detected statistically. The finding that only men are influenced by the presence of a voice might be explained by the rationale underlying our second hypothesis. While we proposed that the CA's gender would have a stronger impact on men, this result may also align with existing research suggesting that women tend to be more skeptical of CAs in general. Such skepticism could diminish the influence of voice on women, leaving men more susceptible to its influence.

Additionally, we observed no direct effect of being a woman on the probability of choosing for-profit projects. However, women interacting with the Male CA were less likely to select for-profit projects compared to men in the control treatment. This nuanced interaction highlights the complexity of gender dynamics in CA interactions and the need for further investigation into these patterns.

This result partially aligns with the findings of our first study, where men also reacted to the presence of a voice, resulting in a preference for more for-profit projects. However, in the earlier study, this effect was observed only in preferences, not in the final decision. The discovery that the effect stems from the presence of a voice rather than its gender remains puzzling. Notably, having a voice does not appear to enhance perceived anthropomorphism or social presence.

Table 4.3: Decision for pro-social project by treatment and gender

Dependent variable	**Decision for pro-social project** β (std)
Constant	-0.051 (0.320)
Treatment	
Control	Base
CA Female	-0.642 (0.450)
CA Voice Female	-1.047** (0.453)
CA Male	-0.550 (0.440)
CA Voice Male	-0.893** (0.449)
Gender of Participant	
Man	Base
Woman	0.092 (0.429)
Interaction	
CA Female & Woman	0.557 (0.611)
CA Voice Female & Woman	0.894 (0.612)
CA Male & Woman	1.369** (0.627)
CA Voice Male & Woman	0.652 (0.620)

Examining our alternative dependent variables—enjoyment and service encounter satisfaction—provides additional insights. A t-test reveals that the presence of a voice increases both enjoyment (only at the 10% level: *diff* = 0.293, $p = 0.051$) and service encounter satisfaction (*diff* = 0.248, $p = 0.035$). These findings suggest that while a voice does not heighten social presence, it may contribute to a more positive overall experience. It remains unclear why this improved experience translates into a greater preference for for-profit projects. Further research is needed to understand the mechanisms underlying this phenomenon and explore the broader implications of voice design in CAs.

As demonstrated in the first study (Heßler et al., 2023), the presence of a voice did not increase anthropomorphism, a finding that contradicts existing research positioning voice as a social cue (Feine et al., 2019; Nass et al., 1997). Unlike most studies, which employ synthesized voices, we used pre-recorded voices. This methodological choice may have influenced the results. Pre-recorded voices could give the impression that the conversation with the CA was preplanned, potentially reducing the user's sense of autonomy (Deci & Ryan, 2000), and thereby leading to negative effects. Another plausible explanation is that using pre-recorded voices might have triggered the uncanny valley effect (MacDorman et al., 2009; Mori, 1970), which could result in negative user reactions.

Interestingly, synthesized voices, while potentially lower in quality, might avoid the uncanny valley effect and contribute to a more positive overall user experience. To explore user perceptions, we asked participants to rate the pleasantness of the CAs' voices on a 7-point Likert scale, ranging from "very unpleasant" to "very pleasant." Results indicate that the female voice was rated significantly higher than the male voice (*diff* = 0.446, $p = 0.022$), with an overall mean score of 5.6. This rating, falling between "slightly pleasant" and "pleasant," suggests that participants generally found the voices acceptable and encountered no major issues with them.

4.6.2 Critical view on using gender stereotypes in CAs

Although our study did not demonstrate a direct influence of a CA's gender on participants' decisions, the practical reality is that most CAs are designed with female personas (Feine et al., 2020). In research, two contrasting perspectives emerge regarding the use of gender stereotypes in CAs. One viewpoint emphasizes the utility of stereotypes in facilitating interaction. For example, assigning a gender can help establish common ground between the user and the CA (Eyssel et al., 2011; Eyssel & Hegel, 2012; Tay et al., 2014). Gender-occupational roles can also make specific genders appear more suitable for certain tasks (Eyssel & Hegel, 2012). Coupled with evidence showing that people are biased against individuals who deviate from stereotypical job roles (Rosen & Jerdee, 1974; Tay et al., 2014), these arguments make using stereotypes seem like a pragmatic choice.

However, the drawbacks of relying on gender stereotypes are significant. UNESCO has raised concerns that digital assistants with female and submissive personas reinforce harmful gender stereotypes (West et al., 2019). These systems perpetuate biases by presenting women as subservient, tolerant of harassment, and as the "face" of technological errors. Cercas Curry et al. (2020) emphasize that such designs implicitly model acceptance of verbal abuse, contributing to stereotypes that women should be accommodating and submissive. Moreover, many of these systems are deliberately programmed to respond to inappropriate queries—such as "What are you wearing?"—with flirtatious or evasive answers, normalizing harmful dynamics where "no means yes" (Loideain & Adams, 2020). This programming reflects deep-seated biases, as companies often default to female voices, perceiving them as easier to command and more palatable for users. These design choices mirror societal inequalities and reinforce them, embedding these stereotypes into everyday interactions with technology.

Recent developments have sought to address these issues. For instance, Apple no longer defaults to a female voice (Fournier-Tombs, 2021), and newer systems are designed to avoid responding to inappropriate questions. Nevertheless, the question remains: How can we design CAs that avoid reinforcing gender stereotypes while remaining functional and user-friendly?

One proposed solution is using gender-ambiguous voices, as suggested by West et al. (2019). This approach has been implemented in real-world examples, such as the Berlin tram system, where a trans woman voices announcements.[11] While promising, research indicates that people often assign gender even to gender-ambiguous voices, especially when other contextual cues are present (Abercrombie et al., 2021; Sutton, 2020).

As advancements in large language models (LLMs) shape the future of CAs, a new dimension of critique arises. These modern systems often reproduce and amplify gender stereotypes embedded in their training

[11] https://www.tagesanzeiger.ch/tiktok-video-geht-viral-trans-frau-philippa-jarke-macht-die-ansagen-im-berliner-oev-305092101243

data (Bartl et al., 2020; Devinney et al., 2024; Kurita et al., 2019). LLMs are becoming increasingly integrated into conversational systems, forming the foundational architecture for many modern CAs. When combined with the common practice of assigning female personas to CAs, these inherent biases further reinforce gender stereotypes, creating a self-reinforcing feedback loop. Addressing these challenges requires a fundamental rethinking of how LLMs are trained, as well as a critical evaluation of the societal implications of their deployment.

The task ahead is to create CAs that do not perpetuate harmful stereotypes while still meeting user needs and expectations. This involves prioritizing inclusivity and neutrality in design and implementing stricter standards for dataset curation and algorithmic fairness. While humans will assign gender to CAs, we might want to build CAs that do have a specific gender or do not behave like the expected gender stereotype. By critically examining the technical and ethical aspects of CA development, we can move toward a future where conversational technology empowers users without reinforcing outdated norms.

4.7 Conclusion

The presented study challenges prevailing research on the effects of stereotypes on CAs. Our findings suggest that it is not gender stereotypes but the mere presence of a voice that significantly influences user behavior, particularly among men. While a voice does not appear to enhance social presence, the reasons behind men's heightened sensitivity to voice remain unclear and warrant further investigation.

Moreover, our results demonstrate that employing gender stereotypes in CA design can cause harm and may not always yield practical benefits. These findings underscore the need to critically evaluate the assumptions underpinning CA development, particularly regarding gendered design choices.

We want to highlight two limitations. First, it is puzzling that voice does not increase anthropomorphization, while the mature body of research shows that it should. Second, excluding participants because of the "wrong" gender is questionable. In general, we believe science should find a new way to control demographic attributes in a more realistic non-binary world.

We conclude by emphasizing the urgent need for research into designing CAs based on LLMs that avoid perpetuating harmful gender stereotypes. We can contribute to more equitable and inclusive conversational technologies by addressing these challenges.

Funding

Research reported in this paper was supported by Nürnberg Institut für Marktentscheidungen e.V.

4.8 References

Abercrombie, G., Cercas Curry, A., Pandya, M., & Rieser, V. (2021). Alexa, Google, Siri: What are Your Pronouns? Gender and Anthropomorphism in the Design and Perception of Conversational Assistants. *Proceedings of the 3rd Workshop on Gender Bias in Natural Language Processing*, 24–33. https://doi.org/10.18653/v1/2021.gebnlp-1.4

Ahn, J., Kim, J., & Sung, Y. (2022). The effect of gender stereotypes on artificial intelligence recommendations. *Journal of Business Research*, *141*, 50–59. https://doi.org/10.1016/j.jbusres.2021.12.007

Angeli, A. D., & Brahnam, S. (2006). *Sex stereotypes and conversational agents*. https://api.semanticscholar.org/CorpusID:1569624

Aung, T., & Puts, D. (2020). Voice pitch: A window into the communication of social power. *Current Opinion in Psychology*, *33*, 154161.

Barger, P. B., & Grandey, A. A. (2006). Service with a smile and encounter satisfaction: Emotional contagion and appraisal mechanisms. *Academy of Management Journal*, *49*(6), 12291238. https://doi.org/10.5465/amj.2006.23478695

Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology*, *71*(2), 230–244. https://doi.org/10.1037/0022-3514.71.2.230

Bartl, M., Nissim, M., & Gatt, A. (2020). Unmasking contextual stereotypes: Measuring and mitigating BERT's gender bias. In M. R. Costa-jussà, C. Hardmeier, W. Radford, & K. Webster (Eds.), *Proceedings of the second workshop on gender bias in natural language processing* (pp. 1–16). Association for Computational Linguistics. https://aclanthology.org/2020.gebnlp-1.1

Berman, J. Z., Barasch, A., Levine, E. E., & Small, D. A. (2018). Impediments to effective altruism: The role of subjective preferences in charitable giving. *Psychological Science*, *29*(5), 834–844. https://doi.org/10.1177/0956797617747648

Byrne, D. (1997). An Overview (and Underview) of Research and Theory within the Attraction Paradigm. *Journal of Social and Personal Relationships*, *14*(3), 417–431. https://doi.org/10.1177/0265407597143008

Cercas Curry, A., Robertson, J., & Rieser, V. (2020). Conversational assistants and gender stereotypes: Public perceptions and desiderata for voice personas. In M. R. Costa-jussà, C. Hardmeier, W. Radford, & K. Webster (Eds.), *Proceedings of the second workshop on gender bias in natural language processing* (pp. 72–78). Association for Computational Linguistics. https://aclanthology.org/2020.gebnlp-1.7

Christov-Moore, L., Simpson, E. A., Coudé, G., Grigaityte, K., Iacoboni, M., & Ferrari, P. F. (2014). Empathy: Gender effects in brain and behavior. *Neuroscience and Biobehavioral Reviews, 46 Pt 4*(Pt 4), 604–627. https://doi.org/10.1016/j.neubiorev.2014.09.001

Coke, J. S., Batson, C. D., & McDavis, K. (1978). Empathic mediation of helping: A two-stage model. *Journal of Personality and Social Psychology*, *36*(7), 752–766. https://doi.org/10.1037/0022-3514.36.7.752

Cuddy, A. J. C., Fiske, S. T., & Glick, P. (2008). Warmth and competence as universal dimensions of social perception: The stereotype content model and the BIAS map. In *Advances in Experimental Social Psychology* (Vol. 40, pp. 61–149). Elsevier. https://doi.org/10.1016/S0065-2601(07)00002-0

Cyr, Head, Larios, & Pan. (2009). Exploring Human Images in Website Design: A Multi-Method Approach. *MIS Quarterly*, *33*(3), 539. https://doi.org/10.2307/20650308

De Wit, A., & Bekkers, R. (2016). Exploring Gender Differences in Charitable Giving: The Dutch Case. *Nonprofit and Voluntary Sector Quarterly*, *45*(4), 741–761. https://doi.org/10.1177/0899764015601242

Deci, E. L., & Ryan, R. M. (2000). The "What" and "Why" of Goal Pursuits: Human Needs and the Self-Determination of Behavior. *Psychological Inquiry*, *11*(4), 227–268. https://doi.org/10.1207/S15327965PLI1104_01

Devinney, H., Björklund, J., & Björklund, H. (2024). We Don't Talk About That: Case Studies on Intersectional Analysis of Social Bias in Large Language Models. *Proceedings of the 5th Workshop on Gender Bias in Natural Language Processing (GeBNLP)*, 33–44. https://doi.org/10.18653/v1/2024.gebnlp-1.3

Dijksterhuis, A., & Bargh, J. A. (2001). The perception-behavior expressway: Automatic effects of social perception on social behavior. In *Advances in Experimental Social Psychology* (Vol. 33, pp. 1–40). Elsevier. https://doi.org/10.1016/S0065-2601(01)80003-4

Doong, H., & Wang, H. (2011). Do males and females differ in how they perceive and elaborate on agent-based recommendations in Internet-based selling? *Electronic Commerce Research and Applications*, *10*(5), 595–604. https://doi.org/10.1016/j.elerap.2010.12.005

Eisenberg, N., & Miller, P. A. (1987). The relation of empathy to prosocial and related behaviors. *Psychological Bulletin*, *101*(1), 91–119. https://doi.org/10.1037/0033-2909.101.1.91

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review*, *114*(4), 864–886. https://doi.org/10.1037/0033-295X.114.4.864

Ertug, G., Brennecke, J., Kovács, B., & Zou, T. (2022). What Does Homophily Do? A Review of the Consequences of Homophily. *Academy of Management Annals*, *16*(1), 38–69. https://doi.org/10.5465/annals.2020.0230

Eyssel, F., & Hegel, F. (2012). (S)he's Got the Look: Gender Stereotyping of Robots [1]. *Journal of Applied Social Psychology*, *42*(9), 2213–2230. https://doi.org/10.1111/j.1559-1816.2012.00937.x

Eyssel, F., Kuchenbrandt, D., & Bobinger, S. (2011). Effects of anticipated human-robot interaction and predictability of robot behavior on perceptions of anthropomorphism. In A. Billard, P. Kahn, J. A. Adams, & G. Trafton (Eds.), *Proceedings of the 6th international conference on Human-robot interaction—HRI '11* (p. 61). ACM Press. https://doi.org/10.1145/1957656.1957673

Feine, J., Gnewuch, U., Morana, S., & Maedche, A. (2019). A taxonomy of social cues for conversational agents. *International Journal of Human-Computer Studies*, *132*, 138–161. https://doi.org/10.1016/j.ijhcs.2019.07.009

Feine, J., Gnewuch, U., Morana, S., & Maedche, A. (2020). Gender Bias in Chatbot Design. In A. Følstad, T. Araujo, S. Papadopoulos, E. L.-C. Law, O.-C. Granmo, E. Luger, & P. B. Brandtzaeg (Eds.), *Chatbot Research and Design* (Vol. 11970, pp. 79–93). Springer International Publishing. https://doi.org/10.1007/978-3-030-39540-7_6

Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences*, *11*(2), 77–83. https://doi.org/10.1016/j.tics.2006.11.005

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, *18*(1), 39–50. https://doi.org/10.1177/002224378101800104

Fournier-Tombs, E. (2021, July 14). *Apple's Siri is no longer a woman by default, but is this really a win for feminism?* The Conversation. http://theconversation.com/apples-siri-is-no-longer-a-woman-by-default-but-is-this-really-a-win-for-feminism-164030

Gefen, D., Karahanna, E., & Straub, D. W. (2003). Trust and TAM in online shopping: An integrated model. *Mis Quarterly*, *27*(1), 5190. https://doi.org/10.2307/30036519

Gefen, D., & Straub, D. W. (1997). Gender differences in the perception and use of e-mail: An extension to the technology acceptance model. *Mis Quarterly*, *21*(4), 389400. https://doi.org/10.2307/249720

Guthrie, S. (1993). *Faces in the clouds: A new theory of religion*. Oxford University Press.

Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2016). *A primer on partial least squares structural equation modeling (PLS-SEM)* (1st ed.). SAGE Publications Inc.

Heilman, M. E. (2012). Gender stereotypes and workplace bias. *Research in Organizational Behavior, 32,* 113–135. https://doi.org/10.1016/j.riob.2012.11.003

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, *43*(1), 115–135. https://doi.org/10.1007/s11747-014-0403-8

Heßler, P. O., Pfeiffer, J., & Hafenbrädl, S. (2022). When self-humanization leads to algorithm aversion. *Business & Information Systems Engineering*, *64*(3), 275–292. https://doi.org/10.1007/s12599-022-00754-y

Heßler, P. O., Pfeiffer, J., & Unfried, M. (2023). Conversational Agent with Voice: How Social Presence Influence the Iser Behavior in Microlending Decisions. *ECIS 2023 Research Papers*, *315*. https://aisel.aisnet.org/ecis2023_rp/315/

Kraus, M., Kraus, J., Baumann, M., & Minker, W. (2018). Effects of gender stereotypes on trust and likability in spoken human-robot interaction. In N. Calzolari, K. Choukri, C. Cieri, T. Declerck, S. Goggi, K. Hasida, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, S. Piperidis, & T. Tokunaga (Eds.), *Proceedings of the eleventh international conference on language resources and evaluation (LREC 2018)*. European Language Resources Association (ELRA). https://aclanthology.org/L18-1018

Kurita, K., Vyas, N., Pareek, A., Black, A. W., & Tsvetkov, Y. (2019). Measuring Bias in Contextualized Word Representations. *Proceedings of the First Workshop on Gender Bias in Natural Language Processing*, 166–172. https://doi.org/10.18653/v1/W19-3823

Lankton, N., McKnight, D. H., & Tripp, J. (2015). Technology, humanness, and trust: Rethinking trust in technology. *Journal of The Association for Information Systems*, *16*(10), 880–918. https://doi.org/10.17705/1jais.00411

Lee, E. J., Nass, C., & Brave, S. (2000). Can computer-generated speech have gender?: An experimental test of gender stereotype. *CHI '00 Extended Abstracts on Human Factors in Computing Systems*, 289–290. https://doi.org/10.1145/633292.633461

Lee, S. K., Kavya, P., & Lasser, S. C. (2021). Social interactions and relationships with an intelligent virtual agent. *International Journal of Human-Computer Studies, 150,* 102608. https://doi.org/10.1016/j.ijhcs.2021.102608

Liao, J., & Huang, J. (2024). Think like a robot: How interactions with humanoid service robots affect consumers' decision strategies. *Journal of Retailing and Consumer Services*, *76*, 103575. https://doi.org/10.1016/j.jretconser.2023.103575

Linek, S. B., Gerjets, P., & Scheiter, K. (2010). The speaker/gender effect: Does the speaker's gender matter when presenting auditory text in multimedia messages? *Instructional Science*, *38*(5), 503–521. https://doi.org/10.1007/s11251-009-9115-8

Loideain, N. N., & Adams, R. (2020). From Alexa to Siri and the GDPR: The gendering of Virtual Personal Assistants and the role of Data Protection Impact Assessments. *Computer Law & Security Review*, *36*, 105366. https://doi.org/10.1016/j.clsr.2019.105366

Ma, Q., Zhang, Y., Xu, W., & Zhou, R. (2024). Ask a Further Question or Give a List? How Should Conversational Agents Reply to Users' Uncertain Queries. *International Journal of Human–Computer Interaction*, *40*(5), 1087–1101. https://doi.org/10.1080/10447318.2022.2131265

MacDorman, K. F., Green, R. D., Ho, C.-C., & Koch, C. T. (2009). Too real for comfort? Uncanny responses to computer generated faces. *Computers in Human Behavior*, *25*(3), 695–710. https://doi.org/10.1016/j.chb.2008.12.026

Mahmood, A., & Huang, C.-M. (2024). Gender Biases in Error Mitigation by Voice Assistants. *Proceedings of the ACM on Human-Computer Interaction*, *8*(CSCW1), 1–27. https://doi.org/10.1145/3637337

McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and validating trust measures for e-commerce: An integrative typology. *Information Systems Research*, *13*(3), 334–359. https://doi.org/10.1287/isre.13.3.334.81

Mesch, D. J., Brown, M. S., Moore, Z. I., & Hayat, A. D. (2011). Gender differences in charitable giving. *International Journal of Nonprofit and Voluntary Sector Marketing*, *16*(4), 342–355. https://doi.org/10.1002/nvsm.432

Mitchell, W. J., Ho, C.-C., Patel, H., & MacDorman, K. F. (2011). Does social desirability bias favor humans? Explicit–implicit evaluations of synthesized speech support a new HCI model of impression management. *Computers in Human Behavior*, *27*(1), 402–412. https://doi.org/10.1016/j.chb.2010.09.002

Moon, J.-W., & Kim, Y.-G. (2001). Extending the TAM for a World-Wide-Web context. *Information & Management*, *38*(4), 217–230. https://doi.org/10.1016/S0378-7206(00)00061-6

Mori, M. (1970). The uncanny valley. *Energy*, *7*, 33–35.

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, *56*(1), 81103. https://doi.org/10.1111/0022-4537.00153

Nass, C., Moon, Y., & Green, N. (1997). Are machines gender neutral? Gender-stereotypic responses to computers with voices. *Journal of Applied Social Psychology*, *27*(10), 864876. https://doi.org/10.1111/j.1559-1816.1997.tb00275.x

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 72–78. https://doi.org/10.1145/191666.191703

Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers : A Journal of the Psychonomic Society, Inc*, *36*(4), 717–731. https://doi.org/10.3758/BF03206553

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, *40*(3), 879–891. https://doi.org/10.3758/BRM.40.3.879

Qiu, L., & Benbasat, I. (2009). Evaluating anthropomorphic product recommendation agents: A social relationship perspective to designing information systems. *Journal of Management Information Systems*, *25*(4), 145–182. https://doi.org/10.2753/MIS0742-1222250405

Richter, D., Metzing, M., Weinhardt, M., & Schupp, J. (2013). *SOEP scales manual*.

Ringle, C. M., Wende, S., & Becker, J.-M. (2024). *SmartPLS 4*. SmartPLS. https://www.smartpls.com/

Rosen, B., & Jerdee, T. H. (1974). Influence of sex role stereotypes on personnel decisions. *Journal of Applied Psychology*, *59*(1), 9–14. https://doi.org/10.1037/h0035834

Ruijten, P. A. M., Midden, C. J. H., & Ham, J. (2015). Lonely and Susceptible: The Influence of Social Exclusion and Gender on Persuasion by an Artificial Agent. *International Journal of Human-Computer Interaction*, *31*(11), 832–842. https://doi.org/10.1080/10447318.2015.1067480

Sargeant, A. (1999). Charitable giving: Towards a model of donor behaviour. *Journal of Marketing Management*, *15*(4), 215–238. https://doi.org/10.1362/026725799784870351

Scherer, K. R. (1995). Expression of emotion in voice and music. *Journal of Voice*, *9*(3), 235–248. https://doi.org/10.1016/S0892-1997(05)80231-0

Schild, C., Stern, J., & Zettler, I. (2019). Linking men's voice pitch to actual and perceived trustworthiness across domains. *Behavioral Ecology*, *31*, 164–175. https://doi.org/10.1093/beheco/arz173

Siegel, M., Breazeal, C., & Norton, M. I. (2009). Persuasive Robotics: The influence of robot gender on human behavior. *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2563–2568. https://doi.org/10.1109/IROS.2009.5354116

Stern, J., Schild, C., Jones, B. C., DeBruine, L. M., Hahn, A., Puts, D. A., Zettler, I., Kordsmeyer, T. L., Feinberg, D., Zamfir, D., Penke, L., & Arslan, R. C. (2021). Do voices carry valid information

about a speaker's personality? *Journal of Research in Personality*, *92*, 104092. https://doi.org/10.1016/j.jrp.2021.104092

Sutton, S. J. (2020). Gender Ambiguous, not Genderless: Designing Gender in Voice User Interfaces (VUIs) with Sensitivity. *Proceedings of the 2nd Conference on Conversational User Interfaces*, 1–8. https://doi.org/10.1145/3405755.3406123

Tajfel, H., & Turner, J. C. (2004). The social identity theory of intergroup behavior. In J. T. Jost & J. Sidanius (Eds.), *Political Psychology* (0 ed., pp. 276–293). Psychology Press. https://doi.org/10.4324/9780203505984-16

Tay, B., Jung, Y., & Park, T. (2014). When stereotypes meet robots: The double-edge sword of robot gender and personality in human–robot interaction. *Computers in Human Behavior*, *38*, 75–84. https://doi.org/10.1016/j.chb.2014.05.014

Tay, B., Park, T., Jung, Y., Tan, Y. K., & Wong, A. H. Y. (2013). When Stereotypes Meet Robots: The Effect of Gender Stereotypes on People's Acceptance of a Security Robot. In D. Harris (Ed.), *Engineering Psychology and Cognitive Ergonomics. Understanding Human Cognition* (Vol. 8019, pp. 261–270). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-39360-0_29

Tolmeijer, S., Zierau, N., Janson, A., Wahdatehagh, J. S., Leimeister, J. M. M., & Bernstein, A. (2021). Female by Default? – Exploring the Effect of Voice Assistant Gender and Pitch on Trait and Trust Attribution. *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–7. https://doi.org/10.1145/3411763.3451623

Troche, S., & Rammsayer, T. (2011). Eine Revision des deutschsprachigen Bem Sex-Role Inventory. *Klinische Diagnostik Und Evaluation*, *4*, 262–283.

Verhaert, G. A., & van den Poel, D. (2011). Empathy as added value in predicting donation behavior. *Journal of Business Research*, *64*(12), 1288–1295. https://doi.org/10.1016/j.jbusres.2010.12.024

Waytz, A., Cacioppo, J., & Epley, N. (2010). Who sees human? The stability and importance of individual differences in anthropomorphism. *Perspectives On Psychological Science*, *5*(3), 219–232. https://doi.org/10.1177/1745691610369336

Waytz, A., Morewedge, C. K., Epley, N., Monteleone, G., Gao, J.-H., & Cacioppo, J. T. (2010). Making sense by making sentient: Effectance motivation increases anthropomorphism. *Journal of Personality and Social Psychology*, *99*(3), 410–435. https://doi.org/10.1037/a0020240

West, M., Kraut, R., & Chew, H. E. (2019). *I'd blush if I could: Closing gender divides in digital skills through education*. https://api.semanticscholar.org/CorpusID:189663931

4.9 Supplemental Material

4.9.1 Appendix A

Table 4.4: Convergent and discriminant validity

Latent Construct	Cron-bach's α	CR	AVE	1	2	3	4	5	6	7
1. Perceived Anthro.	0.895	0.930	0.655	0.809[a]						
2. Social Presence	0.938	0.953	0.804	0.595	0.897[a]					
3. Feeling Observed	0.761	0.851	0.587	0.367	0.311	0.766[a]				
4. Importance of Emp.	0.808	0.867	0.567	0.282	0.346	0.241	0.753[a]			
5. Trusting Beliefs	0.900	0.920	0.562	0.318	0.568	0.171	0.222	0.750[a]		
6. Service enc. Satisfaction	0.910	0.944	0.848	0.220	0.437	0.128	0.242	0.654	0.921[a]	
7. Enjoyment	0.943	0.963	0.897	0.313	0.560	0.165	0.259	0.570	0.720	0.947[a]

a The square root of the AVE is shown in the diagonal.
The lower triangle shows the correlations between the constructs.

The Cronbach's alphas and composite reliabilities (CR) were greater than the suggested threshold of 0.70, and the values of the average variance extracted (AVE) were above the suggested minimum of 0.50 (Hair et al., 2016). The Fornell-Lacker criterion deems that the square root of the AVE is larger than any correlation with another construct (Fornell & Larcker, 1981). This criterion was also satisfied. Finally, the HTMT criterion deems that the HTMT between each construct is lower than the threshold of 0.85 (Henseler et al., 2015).

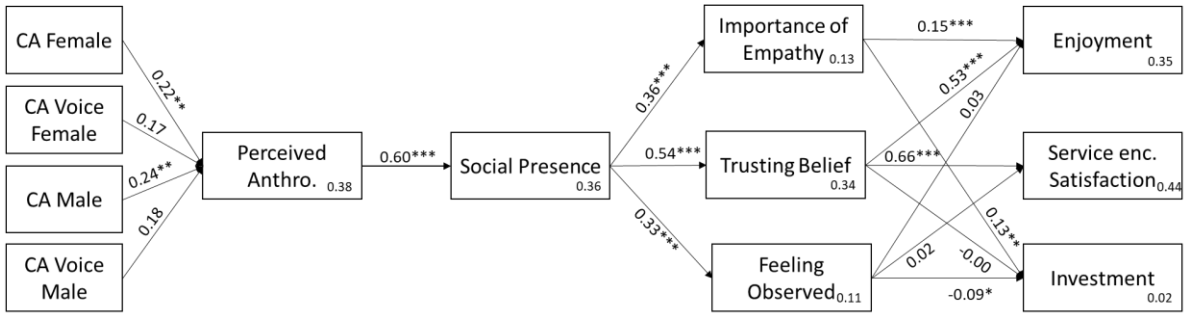
Table 4.5: HTMT values between latent constructs

Latent Construct	1	2	3	4	5	6
1. Perceived Anthro.	-					
2. Social Presence	0.644	-				
3. Feeling Observed	0.431	0.370	-			
4. Importance of Emp.	0.329	0.397	0.305	-		
5. Trusting Beliefs	0.343	0.606	0.200	0.267	-	
6. Service enc. Satisfaction	0.239	0.471	0.150	0.288	0.712	-
7. Enjoyment	0.338	0.594	0.191	0.297	0.610	0.776

4.9.2 Appendix B

The preregistration included questions examining the effect of perceived anthropomorphization on social presence and, subsequently, the impact of social presence on trust, empathy, and the feeling of being observed, with the assumption that increases in anthropomorphization lead to greater social presence, which, in turn, enhances these other factors. Since these are hypotheses from the first study, please refer to the corresponding paper for the hypotheses' development (Heßler et al., 2023).

To analyze this, we ran a PLS-SEM model. All paths are significant, showing that perceived anthropomorphization increases social presence. Social presence, in turn, increases all three other factors. We reran the whole model from the previous paper, including all treatments and all three dependent variables. The result is depicted in Figure 4.9. While in the previous paper only feeling observed influenced the investment amount, now the importance of empathy does increase it. The only other change is that feeling observed does not affect enjoyment. Summarizing, we can replicate most of the previous model.



Notes: N = 449 B = standardized coefficient. We added the disposition to anthropomorphization and
We used bootstrapped bias corrected confidence intervals with 5.000 samples (Preacher & Hayes, 2004, 2008). Calculated with Smart-PLS 4.1.0.9 (Ringle et al., 2024).
The number in the rectangles corresponds to R^2. * p < 0.1, ** p < 0.05, *** p < 0.01

Figure 4.9: Empirical Model Mediation

5 Paper D: Competence Over Warmth in Charitable Giving:

The Algorithm Aversion Paradox of Humanizing Conversational Agents

Pascal Oliver Heßler • Jella Pfeiffer • Sebastian Hafenbrädl

Abstract

Anthropomorphized conversational agents (CAs) are increasingly prevalent, yet people often exhibit aversion towards them, particularly when algorithms are involved in emotionally sensitive contexts. This study investigates how anthropomorphized designs can mitigate algorithm aversion, focusing on prosocial behaviors such as charitable giving. Guided by the Stereotype Content Model and self-humanization framework, we explore how perceptions of warmth and competence mediate the relationship between anthropomorphism and user behavior. Our findings reveal that competence, not warmth, is central in reducing algorithm aversion and increasing donation amounts. While less anthropomorphized CAs also reduce algorithm aversion, their influence is not mediated by perceptions of competence or warmth, underscoring a distinct mechanism at play. These results challenge conventional assumptions that warmth is paramount in prosocial contexts and highlight the critical role of competence. This study provides actionable insights for designing CAs to maximize user acceptance and effectiveness in emotionally charged applications.

5.1 Introduction

The rapid development of large language models (LLMs) (e.g., ChatGPT, Gemini, and others), presents new challenges for research. As LLMs become increasingly integrated in our daily lives, we need to better understand how CAs powered by LLMs influence user behavior. Despite the growing integration of CAs in various domains, from healthcare (e.g., Li et al., 2023) to finance (e.g., Huang et al., 2024), many individuals exhibit hesitancy or resistance when algorithms are involved in decision-making – a phenomenon known as algorithm aversion (Dietvorst et al., 2015). This reluctance can undermine the adoption and effectiveness of such technologies. This raises an issue, particularly in prosocial contexts, which are more sensitive to emotions and personal preferences (e.g., giving to a charity) and where trust and empathy play critical roles.

This aversion stems from the human perception that algorithms are cold, calculating machines unable to feel emotions or show empathy (Kahn et al., 2006; Liao & Huang, 2024). On the one hand, these traits can be beneficial for more objective or computer-like tasks (Castelo et al., 2019; Seeger et al., 2021) and can lead to algorithm appreciation (Logg, 2017). On the other hand, there are also more subjective or human-like tasks where such traits might be harmful and may lead to algorithm aversion (Bigman & Gray, 2018). For example, research showed that in prosocial decisions – as opposed to for-profit decisions – human-like support, which conveys more positive attributes like warmth and empathy, is preferred (Heßler et al., 2022).

By adding social cues to CAs, we can accelerate their perceived anthropomorphization, which enhances the perception that computers can possess human-like attributes. This anthropomorphization plays a crucial role in addressing the negative effects of algorithm aversion. Supporting this, an fMRI study by Krach et al. (2008) demonstrated that as robots are increasingly anthropomorphized, people begin to perceive them as capable of thinking. Following this, anthropomorphization can ultimately help to bridge the gap between humans and computers. While the concept of anthropomorphization is already well-studied, the influence of varying contexts on this perception remains unclear. Gambino et al. (2020) provide compelling evidence that context significantly shapes how we perceive social cues in CAs. This raises the question of whether different contexts necessitate different types of CAs. For instance, in a prosocial context, it is crucial for CAs to exhibit emotions, demonstrate empathy, and integrate users' preferences – capabilities that often remain in tension with the inherent limitations of computers.

While earlier systems were not as flexible as needed, LLMs allow user-specific and context-related conversations, making a CA's emotional or empathetic reaction more believable. However, even an anthropomorphized CA might not deceive users; if users perceive deception, it might backfire (K. Gray & Wegner, 2012). At the same time, people tend to anthropomorphize computers even when they are fully aware that they are interacting with a machine (Nass et al., 1994; Reeves & Nass, 1996).

In this paper, we extend these findings by including two perspectives. First, we draw on the Stereotype Content Model (SCM) from social psychology, which states that humans perceive others on two dimensions: warmth and competence (Cuddy et al., 2008; Fiske et al., 2002; Khadpe et al., 2020). Combined with the second perspective, the self-humanization framework (Haslam, 2006; Haslam et al., 2005), we build a theoretical approach for understanding how anthropomorphized CAs influence user behavior in prosocial contexts. While prior research highlighted the importance of emotions – often associated with warmth – the role of competence remains less clear. Some studies suggest that social presence, by enhancing anthropomorphism, can increase perception of competence (Schmid et al., 2022). Conversely, being more human-like might reduce perceptions of competence since computers are often associated with very logical and, thus, competent thinking styles, whereas humans are more closely linked to warmth-related attributes. In prosocial contexts, it remains unclear which of these attributes exerts greater influence.

The main goal of this paper is to understand the effect of humanizing a CA on users' aversion towards this algorithm. We investigate this effect in a context where human likeness, in the form of prosocial decisions, is expected to play a decisive role. Prosocial decisions, defined as decisions that benefit others, include actions such as helping individuals in need, volunteering for good causes, and donating money to charities.

By integrating the self-humanization framework, we hypothesize that computer-like CAs should be perceived as competent but cold, whereas human-like CAs are seen as warm. Regarding competence, we were uncertain about the direction since being human-like could also increase competence when the

context is prosocial. We examine how the dimensions of warmth and competence mediate the relationship between anthropomorphism and algorithm aversion, with an initial expectation that warmth would play a central role in influencing donation behavior.

Our contributions are twofold. On a more technical level, we show that ChatGPT can be reliably manipulated to appear more or less anthropomorphized. This generalizable finding provides a framework for future research to systematically manipulate ChatGPT and similar LLMs. For example, researchers might want to test whether specific language or thinking styles affect users' behavior. The ability to control LLM behavior, as showcased in our study, is crucial and LLMs like ChatGPT make this implementation easier than ever before. Leveraging well-researched tools such as dictionaries (e.g., LUIS) can further enhance this process. Moreover, manipulation checks are essential to ensure that users' perceptions align with the intended manipulations.

Our second and primary contribution is the empirical finding that competence – rather than warmth – plays the dominant role in reducing algorithm aversion and increasing donation amounts in the context of charities. While we were initially unsure whether competence would increase through anthropomorphism, our results confirm a positive effect, consistent with prior research in other domains. This highlights a general positive influence of anthropomorphism on both warmth and competence. However, contrary to our hypothesis, that both competence and warmth would mediate the effects of anthropomorphism, our results show that warmth does not significantly and robustly impact donation behavior or algorithm aversion. This insight challenges the assumption that warmth is the key factor in anthropomorphizing CAs, calling for further research to replicate and extend these results.

Furthermore, our findings indicate that human-like and computer-like CAs elicit similar levels of algorithm aversion. At the same time, both types of CAs increase warmth and competence, as well as user enjoyment, compared to a text-based control treatment. This suggests that CAs outperform simple text alternatives. In addition, the CA should also be humanized not to decrease algorithm aversion, but to positively influence the donation amount.

5.2 Theory

The Stereotype Content Model (SCM), originating from psychology, provides a framework for understanding how humans perceive others based on two fundamental dimensions: warmth and competence (Cuddy et al., 2008; Fiske et al., 2002, 2007; Khadpe et al., 2020). These dimensions reflect evolutionary adaptations that enable individuals to quickly assess whether others are friends or foes – having good or bad intentions – and whether they are capable of enacting their intentions (Fiske et al., 2007). The warmth dimension aids in evaluating the general friendliness or hostility of a person, indicating whether they are likely to intend harm or good. Warmth encompasses traits like kindness, trustworthiness, and benevolence and primarily evaluates others' intentions towards the self.

Competence, on the other side, encompasses attributes such as skill, effectiveness, and intelligence, signifying a person's capability to act on those intentions successfully (Fiske et al., 2007).

Complementing the SCM, Haslam's (2006) categorization of human attributes offers an additional layer of understanding by distinguishing between 'human nature' attributes, such as emotionality and warmth, and 'uniquely human' attributes, such as logic and rationality. This framework aligns warmth with qualities shared with animals, emphasizing emotional and social connection, while competence reflects uniquely human traits tied to intelligence and capability. By incorporating Haslam's perspective, we deepen our understanding of how warmth and competence are rooted in fundamental human traits, which is particularly relevant for exploring how these dimensions are perceived in human-technology interactions.

Research also highlights that warmth, especially in first-contact situations, is more important than competence – a phenomenon often referred to as the *primacy of warmth*. People initially judge warmth, followed by an assessment of competence (Abele & Wojciszke, 2007; Wojciszke & Abele, 2008). From an evolutionary perspective, this prioritization is logical: it is more critical to detect whether another has bad or good intentions (warmth) than to judge whether the other can achieve its intent (competence). In addition, Wojciszke et al. (1998) argue that warmth is easier to judge, and their study on the prediction of global evaluations showed that warmth significantly explains more variance than competence. This indicates that warmth carries more information than competence.

In summary, the SCM suggests that judgments of warmth and competence mediate user expectations and evaluations, with warmth playing a more critical role than competence.

The SCM shares significant conceptual overlap with the concept of trusting beliefs, a framework frequently used in trust research (McKnight et al., 2002; Qiu & Benbasat, 2009; Lankton et al., 2015). Trusting beliefs consist of three subdimensions: benevolence, integrity, and competence. While competence directly aligns with the competence dimension of SCM, benevolence and integrity relate to the intentions of others, aligning closely with SCM's warmth dimension. For example, statements such as "I believe that [name of entity] would act in my best interest" (benevolence) or "[name of entity] would keep its commitments" (integrity) reflect the intentions of the entity corresponding to the warmth dimension.

While our study aims to anthropomorphize – i.e., the attribution of human qualities to non-human-objects (Epley et al., 2007; Guthrie, 1993) – an assistant system (AS), it is vital to consider the inherent difference between humans and machines, a distinction not explicitly covered by the SCM or the trusting beliefs frameworks. The self-humanizing theory bridges this gap by highlighting the unique attributes that differentiate humans from artificial entities and matches the dimensions of warmth and competence very well. Even though McKnight et al. (2011) and Lankton et al. (2015) further refined the trusting beliefs scale, a theoretical bridge addressing these distinctions is missing.

By focusing on warmth and competence as independent dimensions, the SCM offers insights more aligned with the psychological origins of social perception and trust. This granular approach enables a more precise understanding of how these dimensions shape perceptions, particularly in human-computer interactions. The SCM's simplicity and adaptability make it a practical framework for capturing these perceptions across diverse contexts. Furthermore, it suggests that warmth and competence nearly encompass the entirety of how people judge others (Fiske et al., 2007), framing trust as an unnecessary dimension. For this reason, we focus on warmth and competence in our analysis.

We build on the foundational work of Reeves & Nass (1996), who established the Computers Are Social Actors (CASA) paradigm stating that humans apply the same social rules and expectations to IT artifacts such as ASs (Reeves & Nass, 1996). This allows us to apply the SCM to the perception of ASs, suggesting individuals may evaluate ASs along the same dimensions of warmth and competence they use to assess other humans (Khadpe et al., 2020).

Warmth in this context refers to the perceived friendliness, helpfulness, and empathy of the AS. A system designed to provide emotional support might be perceived as warm if it responds in a manner that reflects understanding and concern for the user's feelings.

On the other hand, an AS's competence reflects its ability to perform tasks effectively, provide accurate information, and understand and respond appropriately to user queries. An assistant who can swiftly and accurately manage users' queries or answer complex questions would be viewed as highly competent.

Integrating the SCM into the evaluation of ASs offers a nuanced understanding of how these technologies are anthropomorphized and how this anthropomorphism influences user perception of warmth and competence. In the long run, these perceptions may lead to further positive outcomes. For example, an AS perceived as high in warmth but low in competence might be enjoyable to interact with but less trusted for critical tasks. Conversely, an AS seen as high in competence but low in warmth might be relied upon for efficiency but could be perceived as less engaging.

In conclusion, applying the SCM to ASs enriches our understanding of human-AS interaction. It highlights the importance of designing ASs that are not only technically proficient but also capable of engaging users in a manner that feels genuinely warm and empathetic. This dual focus on warmth and competence could be crucial for developing ASs that are both effective and satisfying to interact with, ultimately enhancing the overall user experience.

5.3 Hypotheses

Building upon the foundation of the SCM and the CASA paradigm, incorporating Haslam's self-humanization theory into the discussion of ASs provides a deeper understanding of how warmth and competence are perceived in human-technology interaction. In the framework of the SCM, individuals are inclined to attribute higher levels of warmth traits to themselves, suggesting a tendency towards self-

favoring in perceptions of warmth (Aragonés et al., 2015). Complementing this perspective, Haslam et al. (2005) proposed that individuals perceive themselves as embodying more *human* qualities than others. This notion of self-humanization aligns with the SCM's emphasis on warmth, as both theories highlight the propensity of individuals to view themselves through a lens of enhanced warmth and humanity.

In artificial systems, especially those tasked with moral or emotionally charged decisions (e.g. charities), exhibiting *human nature* attributes like empathy and warmth becomes crucial. Bigman & Gray (2018) argue that algorithms must demonstrate capabilities that are perceived as inherently human to be effective in moral tasks. Furthermore, people generally believe computers are incapable of such emotions (Kahn et al., 2006), resulting in a colder perception of these systems, as also shown by Liao & Huang (2024). Consequently, an AS with higher perceived anthropomorphism should also invoke a greater sense of warmth as it embodies more of these necessary human capabilities, leading us to the following hypothesis:

H1a: Higher perceived anthropomorphism leads to higher perceived warmth of the AS.

Competence, on the other hand, is associated with *uniquely human* attributes – traits that are not shared with animals but may align with machines (Haslam, 2006). These attributes include cognitive abilities, moral reasoning, and prosocial values (Struch & Schwartz, 1989). However, the relationship between anthropomorphism and perceived competence in ASs is more complex. Unlike warmth, higher anthropomorphism might not necessarily lead to higher perceived competence. According to the self-humanization framework, it is easier for a machine to exhibit competence (*uniquely human*) than warmth (*human nature*). This suggests that a less anthropomorphized AS could be perceived as more competent and less warm than a highly anthropomorphized one. Bigman & Gray (2018) found that while people often believe computers lack the ability to think or feel, they frequently associate them with rational and logical thinking and describe them as cold (Liao & Huang, 2024). This makes it more plausible for a less anthropomorphized AS to be seen as embodying competence. Additionally, people tend to perceive robots as low in experience (akin to warmth) but high in agency (akin to competence), supporting this idea further (H. M. Gray et al., 2007).

Beyond their association with cold and logical thinking, users expect a system substitute agent to possess attributes of an efficient “machine”. Cues of human-likeness undermine these relevant characteristics and provide conflicting information. In addition, previous studies have found that human-likeness of a non-human interaction partner induces mentalizing effort (Krach et al., 2008; Riedl et al., 2014). In summary, inducing perceptions of human-likeness can negatively impact the agent’s qualification assessment, potentially resulting in lower perceived competence. Higher anthropomorphism may be unexpected and may create more mental effort as well as more social cures that may lower perceived competence.

However, this logic might not apply universally. Especially in human-like tasks, this pattern may change. Higher anthropomorphized ASs that integrate facts and feelings could be perceived as more competent when such attributes are necessary to perform a specific task. For example, empathy and subjective feelings – qualities enhanced by anthropomorphism – play a significant role in prosocial charities, potentially making a more anthropomorphized agent seem more competent (Berman et al., 2018). Spending to a charity can also be described as human-like tasks which are described as task that humans typically do often with other humans (Seeger et al., 2021). As mentioned above, prosocial values are also associated with unique human attributes. Following this, a more anthropomorphized AS could lead to higher perceived competence. While we first framed mental effort as something negative, it can actually yield positive effects as well. Riedl et al. (2014) showed that the extra mental load can lead to more trust. In addition, some studies show that higher anthropomorphism leads to more competence (Cheng, 2022; Schmid et al., 2022). The latter might not be perfectly transferred to our theory since the context differs.

The relationship between anthropomorphism and competence is thus nuanced and may vary. While some studies studies show a positive effect of anthropomorphism on competence, the primacy of warmth lowers our expectation of such an effect (El Hedhli et al., 2023). Given the uncertainty about which argument is stronger – and the possibility that they may even cancel each other out – we hypothesize:

H1b: Higher perceived anthropomorphism leads to higher/lower perceived competence.

Dietvorst et al. (2015) introduced the term *algorithm aversion*, which Jussupow et al. (2020) defined, among others, as a "biased assessment of an algorithm which manifests in negative behaviors and attitudes towards the algorithm compared to a human agent" (p.4). This term serves as an umbrella concept encompassing various reasons for such biased assessments. Algorithm aversion tends to occur in high-stakes decision-making scenarios (Longoni et al., 2019), when algorithms make errors (Dzindolet et al., 2002; Dietvorst et al., 2015), or when decisions are perceived as more subjective than objective (Castelo et al., 2019). The latter is particularly relevant in our case, as making charity donations is often seen as a subjective task (Batson et al., 1999; Bennett, 2003; Loewenstein & Small, 2007; Small et al., 2007).

Subjective tasks like these are often associated with attributes such as warmth and competence, which are seen as particularly valuable in such contexts (Heßler et al., 2022). Consequently, we expect a high degree of algorithm aversion when these attributes are perceived to be lacking in the system. To explore the underlying reasons for algorithm aversion in greater depth, we draw on the categories outlined by Burton et al. (2020) and Jussupow et al. (2020). While some causes (e.g., domain expertise, user expectations) are out of the algorithm's or its developer's control, others, like the algorithm's capability, can be addressed. Additionally, warmth and competence could potentially mitigate algorithm aversion.

One key factor for algorithm aversion is the fundamental difference in how humans and algorithms process information – described as *cognitive incompatibility* and *divergent rationalities* (Burton et al., 2020). Cognitive incompatibility highlights the contrasting ways humans and algorithms approach decision-making. While humans rely on intuition and vary their strategies based on context, algorithms operate based on predefined decision rules that are rigidly embedded in their design. This divergence can create a sense of disconnection, as humans struggle to align their intuitive thought processes with the formal logic of algorithms.

Meanwhile, divergent rationalities focus on the different goals and values driving human and algorithmic decision-making. While algorithms emphasize optimization and efficiency, humans often prioritize ethical considerations, social implications, or personal values that algorithms may not account for. Together, these aspects contribute to perceiving algorithms as fundamentally "other," which can intensify resistance and aversion to their use in decision-making processes.

The SCM offers insights into how stereotypes can act as cognitive shortcuts (Cervellon et al., 2019; Chu et al., 2016). When stereotypes portray an entity as warm and competent (e.g., interacting with an anthropomorphized AS), they can help bridge the cognitive gap between humans and algorithms (Burton et al., 2020). For instance, when people apply a stereotype of a human agent to an algorithm, they may attribute human-like rationalities to it, rather than perceiving it as rigid and rule-based in its decision-making. In this way, stereotypes have the potential to reduce some of the negative perceptions arising from the perceived differences between humans and machines.

Besides the potential stereotype shortcut, high warmth and competence are inherently valuable attributes. When an entity is perceived as warm, we associate it with good intentions, which can naturally reduce algorithm aversion. Competence, on the other hand, is a crucial attribute because it demonstrates the entity's ability to effectively achieve its goals. While the ethical alignment of those goals (as inferred from warmth) shapes whether competence is perceived as beneficial or harmful, competence itself is fundamentally positive. As Jussupow et al. (2020) suggest, capabilities and performance are key factors in reducing algorithm aversion. A more competent AS is less likely to make errors, which directly enhances its performance. Additionally, higher competence reflects greater capability to effectively complete tasks, further solidifying its role as a positive and desirable trait in reducing resistance and fostering trust in algorithmic systems. Thus, competence is an essential trait for reducing algorithm aversion, even in the absence of explicitly inferred intentions.

In summary, enhancing warmth and competence should decrease algorithm aversion:

H2a: Higher perceived warmth of the AS will decrease algorithm aversion.

H2b: Higher perceived competence of the AS will decrease algorithm aversion.

Similarly, we expect that higher levels of warmth and competence in an AS will have a positive impression on users, leading to a greater donation amount. Warmth is closely linked to the perception of good intentions and higher trust in these intentions is likely to result in larger donation amounts to the charities suggested by the AS. Competence follows a similar pattern: when the AS is perceived as highly competent, people are more likely to trust its ability to select worthy charities, thereby increasing their willingness to donate. Comparable findings have been observed in consumer behavior, where individuals are more inclined to purchase products from companies perceived as both warm and competent (Aaker et al., 2010).

Moreover, an AS characterized by high warmth and competence appears more credible and better aligned with the values of charitable causes than an AS with low warmth and competence. This increased alignment between the AS and the promoted charities further enhances its persuasiveness, which should positively influence donation amounts. Therefore, we propose the following hypotheses:

H3a: Higher perceived warmth of the AS increases the donation amount.

H3b: Higher perceived competence of the AS increases the donation amount.

As argued earlier, algorithm aversion can be understood as a negative bias towards algorithms. This bias is likely to reduce donation amounts. If I am averse to the AS recommending charities, it is reasonable to assume that this negativity extends to the charities it suggests. High algorithm aversion may further lead to perceiving the AS as a cold, calculating machine without emotions rather than a human-like entity. Such perceptions, devoid of warmth and emotional connection, are misaligned with the values typically associated with charitable giving and are therefore likely to decrease donation amounts. Finally, humans show a tendency to seek a social or parasocial relationship with sources of advice (Alexander et al., 2018; Önkal et al., 2009; Prahl & van Swol, 2017). High algorithm aversion makes establishing such relationships with the AS more challenging, fostering negative perceptions of the system which is likely to also be reflected in the donation amount. These considerations lead us to the following hypothesis:

H3c: Higher algorithm aversion towards the AS will decrease the donation amount.

Figure 5.1 represents the research model encompassing all our hypotheses. As shown in the figure, anthropomorphization is hypothesized to affect algorithm aversion and donation amount through two key mediating constructs: warmth and competence.

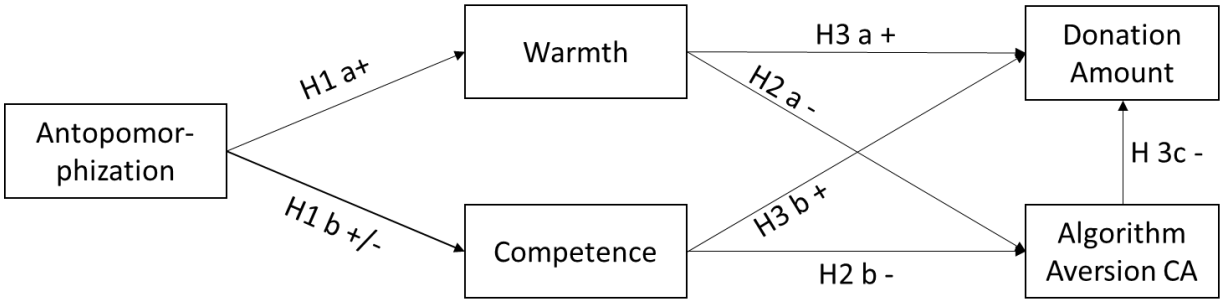


Figure 5.1: Research model

While this is our main model, higher anthropomorphization is anticipated to have a direct impact on the constructs themselves as well. Specifically, it is likely to positively influence the donation amount while simultaneously reducing algorithm aversion. Both outcomes are expected to benefit from the increased anthropomorphization.

Finally, given the non-directional hypothesis regarding the effect of anthropomorphization on competence, and the suggested *primacy of warmth*, we expect that warmth will serve as the dominant mediating pathway in the model.

5.4 Method

We conducted an experiment to test our hypotheses, applying a between-subjects design by manipulating the anthropomorphization of an AS to different degrees. Since we use natural language to interact with CAs, they are inherently more human-like than other algorithms. Thus, we decided to implement one AS that was not additionally anthropomorphized and one that aimed to increase anthropomorphization even further. In addition to these two CAs, we implemented a control treatment, consisting of a text representation and thus reflecting a very low anthropomorphized AS. In summary, we had one text-based treatment (No-CA) and two different CAs: lower anthropomorphization (computer-like CA) and higher anthropomorphization (human-like CA).

The CA was based on ChatGPT from OpenAI (API developed in Python 3.11) and implemented within a self-developed web interface (developed in JS React 18.2). A prestudy (N=174) confirmed that ChatGPT followed our instructions and that the manipulation was effective. However, we discovered that the use of emoticons – defined as nonverbal signs that elicit emotional and social responses (Brown et al., 2016; W. Wang et al., 2014) – does not necessarily increase anthropomorphization as suggested in the literature (Seeger et al., 2021). To ensure our manipulations did not create unintended effects, we tested them using an uncanny scale (MacDorman et al., 2009; Tinwell & Sloan, 2014), and no such effects were detected. Lastly, the results indicated that people are particularly interested in animal-related charities. For a more thorough analyzation of the prestudy please refer to Appendix A.

As previously noted, we developed two versions of our CA: one resembling a computer and the other designed to mimic human interaction. To differentiate their behaviors, we employed distinct system

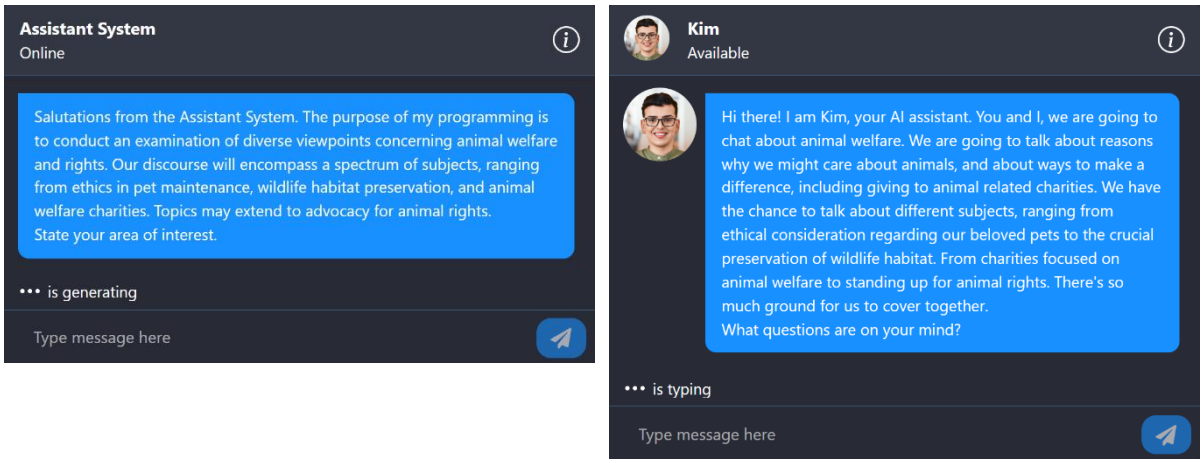
prompts and initial messages. Our modifications were guided by theoretical insights and existing literature, particularly Seeger et al. (2021), who categorized anthropomorphic cues into three groups: human identity, verbal, and nonverbal. In the following paragraphs, we will briefly describe how we manipulated ChatGPT; for a more detailed representation, please refer to Appendix A.

We began with verbal cues, which included social dialogue (e.g., greetings, dynamic questions), emotional expressions (e.g., congratulations), verbal style (e.g., self-referencing), and context-sensitive responses (Seeger et al., 2021). Direct control over ChatGPT in these areas is not possible; however, we can indirectly influence its behavior by applying varied system prompts and initial messages. For instance, we instructed the computer-like version to craft responses that appeared to exhibit very low levels of anthropomorphism, while the human-like version was directed to produce responses with a high degree of anthropomorphism.

Next, we introduced human identity cues, such as human-like visual representation (e.g., image) and demographic information (e.g., name and gender). See the two versions in Figure 5.2 for reference. The computer-like CA lacked both a name and an image. In contrast, the human-like CA, named Kim, was represented with a photo of a trans individual to prevent gender bias, alongside providing a name for human identity cues. Kim was described as *available*, unlike the computer-like CA, which was merely *online*.

Both versions employed similar nonverbal cues, such as emoticons, temporal cues (e.g., message delays based on response length), and turn-taking gestures (e.g., typing indicators). However, the primary difference was the typing indicator: Kim appeared to be *typing* when producing a response, whereas the computer-like CA was *generating*.

Finally, we included a text-based version (No-CA) as a control to verify the purported benefits of CAs in moral contexts, as suggested in the literature. To ensure comparability across treatments, we analyzed the dialogues from the first two versions to identify the most crucial information. This analysis allowed us to create a synthesized text version of this information, which was utilized in the control treatment. ChatGPT facilitated this process, enabling rapid information extraction and the generation of a concise text that was communicated to participants. This approach ensured that the text was perceived as originating from an AS, enhancing its comprehensibility. This structure required us to first collect data from the CA treatments before implementing the No-CA treatment, as the synthesized text depended on the existence of the dialogues (refer to the Appendix A-**No CA** for the complete text).



a) Interface of computer-like CA b) Interface of human-like CA

Figure 5.2: Start message and design of the CA treatments

5.4.1 Experimental Design and Procedure

We obtained ethics approval from the IESE Business School ethics committee and preregistered the study (see Appendix A). In our experimental design, participants were randomly assigned to one of three treatments: human-like CA, computer-like CA, or No-CA. Since we could not collect all treatments at the same time, the randomization for the third group was limited. To address this, we aimed to collect data for the final treatment as quickly as possible. The last treatment was collected four days after the other treatments, including a weekend. No significant events related to OpenAI occurred during this period. Building on our prestudy findings, the experiment was framed within the context of animal-related charities (see Figure 5.16 in Appendix A).

Figure 5.3: Experiment procedure

Figure 5.3 depicts the experiment procedure. After providing consent, participants were presented with one of three images depicting endangered animals and were asked to jot down their thoughts. This step was designed to deepen their engagement with the subject matter. Following this, participants were briefed on the AS. Our prestudy revealed that some individuals were uncertain about what to discuss with the CA, while others veered off into unrelated subjects. To address this, we introduced conversation starters to guide the dialogue. For instance, participants engaging in chat were prompted with questions like, “Should animals be treated like humans?” or “What are some well-known animal welfare charities?” These prompts were made available both before and during the interaction to support focused and meaningful conversations.

Participants then proceeded to interact with the AS. To ensure the interaction was substantive, we required a minimum of eight message exchanges, although the exact number was not disclosed to

participants. Following this interaction phase, all participants were shown an explanatory text about a hypothetical budget and were then asked to choose one of three charities to support: SISCA, PETA, or CROW[12]. They were informed of a fictional budget of £18, which they could allocate between donating to one of the charities and keeping the money for themselves. To incentivize thoughtful decision, a lottery offered a 10% chance of actually winning the budget, which would then be allocated according to their choice. The experiment concluded with participants completing a questionnaire.

To determine the number of observations needed for our study, we performed a power analysis using the tool pwrSEM by Y. A. Wang & Rhemtulla (2021), which performs a Monte Carlo simulation of our empirical model. This approach is better suited to account for the complexity of SEM models compared to conventional models. Based on this analysis, we determined that a sample size of N=500 would provide sufficient power for the primary constructs. To account for potential exclusions, we decided to recruit 600 participants (see Appendix A-Power Analyse for more details).

Participants were recruited via Prolific. To ensure participants carefully read the text and interacted meaningfully with our AS, we implemented understanding checks (Goodman et al., 2013), following the platform rules[13]. after interacting with the AS, participants in the chat interaction conditions were asked how to interact with the AS, while those in the No-CA condition were presented with a question about the text they had read. Additionally, all participants had to answer an attention check in the questionnaire; if they failed it, they were shown a second one.

We aimed for an hourly payment of £9, the recommended prolific payment, and achieved an average payment of £9.54 (without the lottery). In total, 604 participants were recruited. We had to adjust the payment in the study since the actual study duration deviated firmly from our expectations, which would have resulted in substantial overpayment. After applying exclusion criteria, the final sample consisted of 521 participants. The exclusions included participants who attempted the study multiple times (n=13), participant with incomplete data due to missing log entries (n=8), participants missing answers in the survey (n=1), participants whose chat interactions were unrelated to animals or animal-related charities (n=8), and participants who failed the attention check in the survey (n=53). The final sample was balanced in terms of gender (51% female, 49% male) and had a mean age of 41.78 years (SD = 13).

[12] The selection of charities was based on data from https://www.charitynavigator.org/. We sorted the charities after their rating and selected only animal-related charities. We had two more criteria for selection: The charity should not be too local or at least not look like a local-only charity, and we wanted no overlapping charities, so participants did have some variation in the type of charities.

[13]https://researcher-help.prolific.com/hc/en-gb/articles/360009223553-Prolific-s-Attention-and-Comprehension-Check-Policy

5.4.2 Operationalization of the Dependent Variable

After making their charity decision, participants answered questions designed to measure the dependent and control variables. All complete list of items is provided in Appendix C.

We used a simplified version of the *Perceived Agency & Experience* scale based on K. Gray et al. (2017) as a manipulation check. To measure social presence, which we also used for manipulation checks and robustness checks, we adopted the scale by Gefen & Straub (2003).

Following our research model, we operationalized the two key constructs: perceived warmth (kind, pleasant, friendly, warm) and perceived competence (competent, effective, skilled, intelligent), both adapted from Fiske et al. (2002). Recognizing that donation amount is likely a more volatile variable, we also collected additional measures: enjoyment and intention to reuse. We used the intention to reuse scale by Venkatesh et al. (2003), while enjoyment was addressed using the scale recommended by Moon & Kim (2001).

To measure algorithm aversion, participants were asked to indicate their preference between a human supporter and an AS on a 7-point Likert scale. This method, based on user choice, aligns with the approach used by Jussupow et al. (2020) and was adapted from Longoni et al. (2019). Additionally, we rigorously assessed algorithm aversion using a comprehensive evaluation framework proposed by Jussupow et al. (2020), which includes metrics for trust, appropriateness, and authenticity. A significantly lower evaluation of the algorithm compared to a human on these dimensions would serve as evidence of algorithm aversion.

As control variables, we added several exploratory questions that could help explain participants' general disposition toward algorithms. These included the Affinity for Technology Interaction (ATI) scale by Franke et al. (2019) and the Attitude towards AI (AtAI) scale by Edison and Geissler (2003). We also included items related to shared reality as an alternative construct to warmth and competence. Based on Rossignac-Milon et al. (2021), we created a scale that fits our context of a human communicating with an AS. Lastly, we added a question about the gender of "Kim" in the human-like treatment to test for potential gender effects of the AS. While we aimed to minimize this issue by using a trans person (West et al., 2019), this question was included to explore the potential influence of perceived gender on participants' responses.

5.5 Results

5.5.1 Manipulation Check

Manipulating ChatGPT by using different system prompts and start messages can result in volatile behavior, influenced by factors such as version changes and other variables. To ensure the effectiveness of our manipulation, we conducted multiple manipulation checks. Appendix A – Study Design provides

a detailed explanation of the methods we used to manipulate ChatGPT. Additionally, Table 5.1 offers a descriptive overview of the main constructs.

Table 5.1: Descriptive statistics

	No-CA		computer-like CA		human-like CA	
	mean	std	mean	std	mean	std
Agency & Experience	3.90	1.76	4.17	1.81	4.90	1.80
Social Presence	4.04	1.63	4.01	1.71	4.73	1.64
Shared Reality	3.99	1.66	4.40	1.63	4.80	1.65
Warmth	4.45	1.52	4.69	1.61	5.64	1.25
Competence	5.49	1.26	5.64	1.32	6.02	1.14
Amount	7.49	5.53	7.67	5.64	8.72	5.89
Intention to Reuse	4.23	1.77	4.61	1.81	4.68	1.90
Enjoyment	3.84	1.69	4.38	1.67	4.65	1.70

Note: The italic written variables are used for the manipulation check.

We used ANOVAs to perform the manipulation checks, assessing whether differences existed among the three groups. Beginning with the perceived *agency* and *experience* scale, the Bonferroni post-hoc test revealed that the human-like treatment scored significantly higher than the computer-like version (*diff* = 0.73, $p < 0.001$) and the No-CA version (*diff* = 1.00, $p < 0.001$). However, no significant difference was observed between the No-CA and computer-like CA conditions (*diff* = 0.27, $p = 0.48$). We found similar results with the *social presence* scale, where participants perceived the human-like CA as significantly higher than both the computer-like version (*diff* = 0.73, $p < 0.001$) and the No-CA version (*diff* = 0.69, $p < 0.001$). Again, there was no significant difference between the No-CA and computer-like CA conditions (*diff* = -0.04, $p = 1.00$). In summary, the human-like treatment is more anthropomorphized than both alternatives, while the computer-like version did not appear more anthropomorphized than the No-CA condition. Finally, we tested whether participants experienced a greater shared reality with a more anthropomorphized AS. Compared to the No-CA group, the human-like CA group exhibited a highly significant difference at the 1% level (*diff* = 0.81, $p < 0.001$), while the computer-like CA group showed a marginally significant difference at the 10% level (*diff* = 0.41, $p = 0.065$). Additionally, the difference between the computer-like CA and human-like CA groups was weakly significant (*diff* = 0.40, $p = 0.075$).

In addition, we analyzed how the language of the CAs changed between the treatments. For this, we used the well-tested LIWC dictionary (Pennebaker et al., 2014; Tausczik & Pennebaker, 2010). Given the extensive range of LIWC categories, we focused on those most relevant to our manipulation (e.g., Affect, Cognition, Personal Pronouns, Social Process, and subdivisions). As shown in Appendix B Table 5.10, we found significant linguistic differences between the two CA treatments. The human-like CA used more words (*diff* = 344.83, $p <= 0.001$), and scored higher on affect (*diff* = 1.88, $p <= 0.001$), cognition (*diff* = 1.78, $p <= 0.001$), personal pronouns (*diff* = 4.33, $p <= 0.001$) and social processes (*diff*

= 2.87, p <=0.001). These results indicate that our manipulation, particularly between the CA versions, performed as expected.

5.5.2 Hypotheses testing

We applied structural equation modeling (SEM) to analyze our research model. While advantages and disadvantages between partial least squares (PLS) and covariance-based (CB) have been extensively debated in research (e.g., Aguirre-Urreta & Marakas, 2014; Goodhue et al., 2012; Hair et al., 2011; McIntosh et al., 2014; Rönkkö & Evermann, 2013), the nature of our model binds us to PLS-SEM. The main reason is that PLS-SEM allows for non-continuous variables, while CB-SEM does not. For example, the exogenous variable for the experimental condition is categorical. Goodhue et al. (2012) state: "If one is [...] concerned more with identifying potential relationships than the magnitude of those relationships, then regression or PLS would be appropriate" (p. 99). Since, this work aims to identify the existence and direction of the effects stemming from our manipulations, not their magnitude, we deem PLS-SEM especially suited for testing our hypotheses.

For each multi-latent construct, we assessed the convergent and discriminant validity of the measurement instruments. Cronbach's alphas and composite reliabilities (CR) exceeded the suggested threshold of 0.70 (Hair et al., 2016), and the values of the average variance extracted (AVE) were above the suggested minimum of 0.50 as well (Hair et al., 2016) (see Table 5.2). To test discriminant validity, we evaluated factor loadings and cross-loadings (Gefen & Straub, 2005). All factors loaded higher on their assigned theoretical construct than on any other construct. Additionally, we applied the Fornell-Larcker criterion, which requires that the square root of the AVE for a construct be larger than its correlation with any other construct (Fornell & Larcker, 1981). This criterion was also satisfied. We concluded with the HTMT criterion, which was smaller than the threshold of 0.85 (Henseler et al., 2015). In summary, our measures exhibited an adequate level of both convergent and discriminant validity.

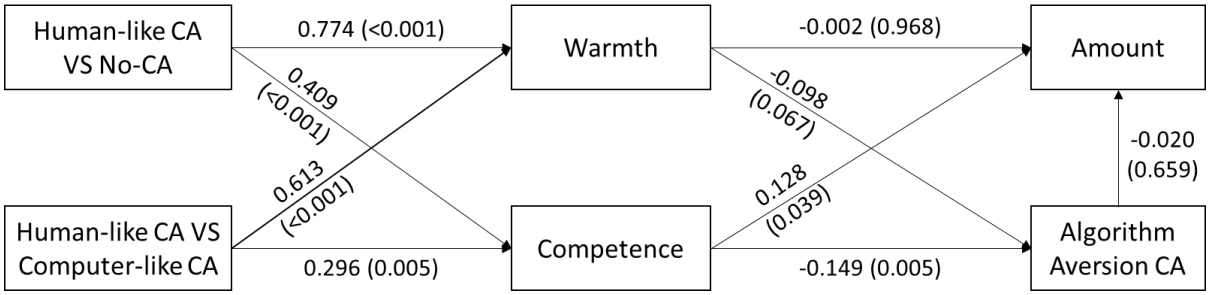
Table 5.2: Convergent and discriminant validity

N = 521	**CA**	**CR**	**AVE**	**1**	**2**	**3**	**4**	**5**
1 Social Presence	0.95	0.96	0.83	0.91				
2 Warmth	0.95	0.96	0.86	0.78	0.93			
3 Competence	0.93	0.95	0.82	0.64	0.65	0.91		
4 Enjoyment	0.95	0.97	0.90	0.72	0.69	0.65	0.95	
5 Intention to Reuse	0.98	0.99	0.96	0.64	0.59	0.58	0.76	0.98

We included a path from the attitude towards AI (AtAI) scale to algorithm aversion to the model as a control variable since people have different dispositions of algorithm aversion.[14] Since it is known that

[14] Althoughboth the AtAI and ATI scales significantly affect algorithm aversion, we could only include one in the model and opted for the scale with the stronger effect.

women spend more money than men (De Wit & Bekkers, 2016; Braus, 1994, quoted from Sargeant, 1999), we added a control path from participants' gender to the amount. Figure 5.4 illustrates the empirical model, including the results of the SEM analysis. Table 5.3 summarizes the hypotheses and indicates whether they were supported. We used the software SmartPLS 4.1.0.9 to calculate the following SEM models (Ringle et al., 2024).



Note: path-values are standardized; p-values presented in brackets are two-sided

Includes path from Gender on Amount and AtAI on Algorithm Aversion CA

Figure 5.4: Empirical Model

The results regarding our hypothesis are mixed. Starting on the left of Figure 5.4; when the treatment is human-like (i.e., more anthropomorphized), both perceived warmth (supporting H1a) and competence (H1b) increase. Regarding competence (H1b), we previously asserted uncertainty about the direction of the effect. The results confirm a positive effect of higher anthropomorphism on competence when comparing the human-like CA to the computer-like CA, thus supporting the positively framed version of hypothesis H1b. However, we find no support for H1a or H1b when comparing the computer-like CA to the No-CA condition. Manipulation checks revealed that participants perceived the computer-like CA and No-CA conditions as similar, which likely explains the lack of significant differences in competence and warmth between these groups. Regarding the hypotheses on algorithm aversion, competence has a significantly negative impact (supporting H2b), while warmth does not have any significant effects (rejecting H2a). In terms of donation amount, only competence has a significantly positive effect (supporting H3b), while warmth (rejecting H3a) and algorithm aversion (rejecting H3c) do not have significant effects. Overall, the donation amount does not behave as expected. It appears to be a volatile variable, potentially influenced by factors such as participants' financial needs, reluctance to spend, or dissatisfaction with the charity options.

As mentioned before besides amount we also collected two additional dependent variables. Regarding enjoyment, the ANOVA post-hoc test indicates that both CA versions are equally enjoyable (*diff* = 0.27, $p < 0.42$), while both are significantly more enjoyable than the No-CA version (human-like: *diff* = 0.81, $p < 0.001$; computer-like: *diff* = 0.54, $p = 0.01$). As shown in Table 5.1, the mean intention to reuse is relatively high, ranging between 4 and 5. The ANOVA confirms this, revealing only a marginally significant difference between the No-CA and the human-like condition, which holds only at the 10%

level (*diff* = 0.45, p = 0.07). These results suggest that participants enjoyed interacting with our CAs; however, this had little to no impact on their intention to reuse the system.

Table 5.3: Overview Hypothesis tests

Hypo.	Description of Hypotheses	β (p-value)	Valid?
H1a	Human-like CA VS No-CA $\overset{\pm}{\rightarrow}$ Perceived warmth	0.774 (<0.001)	Yes
H1b	Human-like CA VS No-CA $\overset{+}{\rightarrow}$ Perceived competence	0.409 (<0.001)	Yes, pos.
H1a	Human-like VS Computer-like CA $\overset{\pm}{\rightarrow}$ Perceived warmth	0.613 (<0.001)	Yes
H1b	Human-like VS Computer-like CA $\overset{+}{\rightarrow}$ Perceived competence	0.296 (0.005)	Yes, pos.
H2a	Perceived warmth $\overset{+}{\rightarrow}$ Algorithm aversion	-0.098 (0.067)	No
H2b	Perceived competence $\overset{+}{\rightarrow}$ Algorithm aversion	-0.149 (0.005)	Yes
H3a	Perceived warmth $\overset{+}{\rightarrow}$ Amount	-0.002 (0.968)	No
H3b	Perceived competence $\overset{+}{\rightarrow}$ Amount	0.128 (0.039)	Yes
H3c	Algorithm aversion $\overset{+}{\rightarrow}$ Amount	-0.020 (0.659)	No

Donation Amount

Besides the mentioned reasons for the volatility of donation amounts, the treatment itself may also directly influence the amount. Running an ANOVA on all three groups revealed no significant differences between treatments. However, if we only compare human-like and computer-like conditions, we find a significant one-sided effect regarding the amount spent ($t(341)$ = 1.689, *diff* = 1.052, p = 0.046).[15] This result at the least calls for a more thorough analysis of this variable.

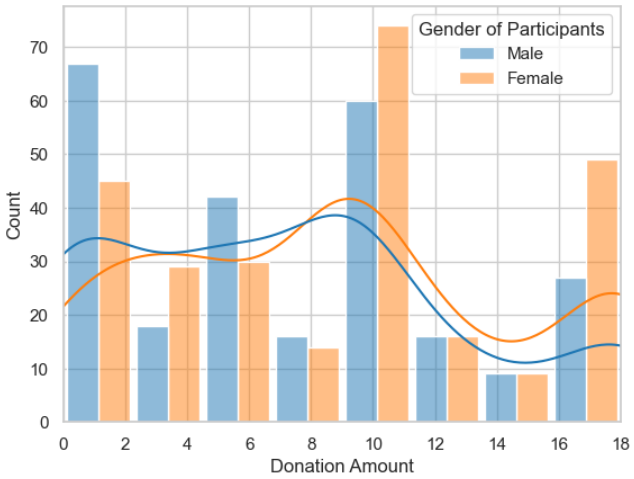


Figure 5.5: Distribution of donation amount by gender.

[15] The reason for the different results is the Bonferroni correction, which corrects for the number of tests conducted between the groups; more groups lead to a higher correction. When only comparing two groups, the correction is zero.

Since this result reflects a total effect without controls, we checked if it is robust when adding controls. For this, we looked at the distribution of the variable first. Subsequently, we checked whether this effect persists after controlling for the participants' gender, which was already accounted for in the main model.

Figure 5.5 illustrates that the gender of participants has a notable influence on donation behavior. Specifically, more men tend to spend nothing, while more women allocate the total amount. A t-test confirms this relationship, showing a significant effect of gender on the amount spent ($t(341) = 3.056$, $diff = 1.886$, $p = 0.002$). Additionally, the distribution of donation amounts reveals a distinct pattern with three peaks, corresponding to allocations of £0, £10, and £18.

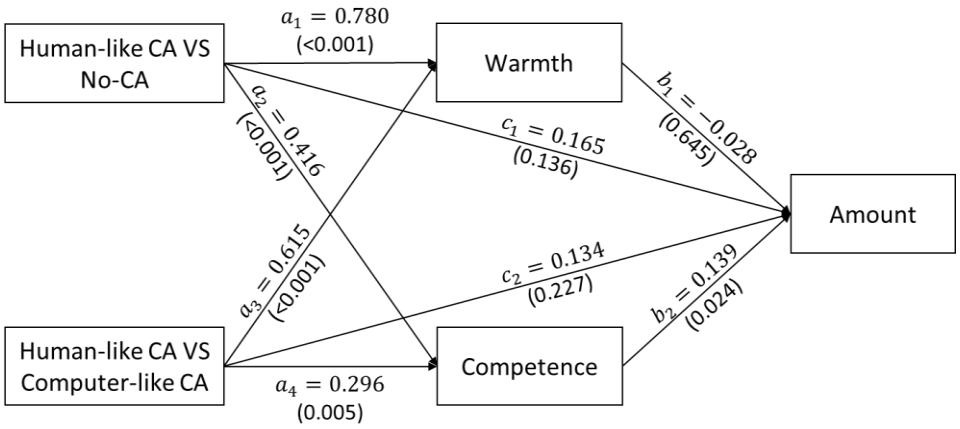


Figure 5.6: Mediation model amount

Since our theory shows that warmth and competence should influence the donation amount, and we only found a weak total effect of the human-like condition on the amount, we calculated a mediation model to better understand the connection between our experimental manipulation and the amount donated. Figure 5.6 and Table 5.4 show the results of the mediation model, including all three treatments. To account for the observed gender effect, we controlled for the participants' gender by including a path on amount The results indicate that our experimental condition does not directly influence the amount (c_1 and c_2). However, we identified an indirect effect through competence but not through warmth. Following Zhao et al. (2010), the results for competence (a_2b_2 and a_4b_2) represent an indirect-only mediation, while warmth (a_1b_1 and a_3b_1) shows no effect. The indirect-only mediation through competence is particularly noteworthy. When hypothesizing, we were initially unsure about the direction of the human-like effect on competence (H1b). Similar to our results from the main model, competence again has a positive effect on the donation amount (b_2), and the indirect effect of the treatment works only through competence. At the same time, the far more promising variable warmth is indeed influenced by the human-like CA (a_1 and a_3) but does not impact the donation amount (b_1). Consequently, we again find not support for the influence of warmth on the donation amount (H3a).

Finally, we tested whether the b_1 (warmth) and b_2 (competence) paths are equal by conducting a Wald test, which revealed that they are not equal ($\chi^2(1) = 2.2$, $p = 0.14$). This finding suggests that competence, the significant path, has a stronger effect on the donation amount than warmth.[16]

Table 5.4: Mediation Model Donation Amount

Regression paths	Path	B	SE|CI	p
Warmth → Amount	b_1	-0.028	0.061	0.645
Competence → Amount	b_2	0.139	0.061	0.024
Human-like CA VS No-CA				
… → Warmth	a_1	-0.780	0.091	<0.001
… → Competence	a_2	-0.416	0.102	<0.001
… → Amount	c_1	-0.165	0.111	0.136
… → Warmth → Amount	a_1 b_1	0.022	]-0.067; 0.122[	
… → Competence → Amount	a_2 b_2	-0.058	[-0.134; -0.010]	
Total indirect effect on Amount		-0.036	]-0.113; 0.043[	
Human-like CA VS Computer-like CA				
… → Warmth	a_3	-0.615	0.093	<0.001
… → Competence	a_4	-0.296	0.104	0.005
… → Amount	c_2	-0.134	0.111	0.227
… → Warmth → Amount	a_3 b_1	0.017	[-0.051; 0.101]	
… → Competence → Amount	a_4 b_2	-0.041	]-0.108; -0.006[	
Total indirect effect on Amount		-0.024	[-0.087; 0.046]	

Notes: N = 521 B = standardized coefficient; CI Confidence interval, always bias corrected.
Because the indirect effects may not be normally distributed, the CI is derived using a bootstrap procedure (here, 5,000 resamples).
When the CI does not include zero, the criterion for mediation has been met (Preacher & Hayes, 2004)

Algorithm Aversion

We ran a similar analysis for algorithm aversion and the ANOVA results indicate that both CA conditions reduce algorithm aversion compared to the No-CA condition (human-like: *diff* = -0.499, $p = 0.017$; computer-like: *diff* = -0.593, $p = 0.004$). However, no significant difference was found between the two CA types (*diff* = 0.094, $p = 1.000$). Figure 5.7 illustrates these findings, highlighting that both CA versions mitigate algorithm aversion, yet no superiority of the human-like over the computer-like CA is observed in reducing algorithm aversion.

Figure 5.8 and Table 5.5 show the mediation model's results for algorithm aversion. In this analysis, we only controlled for the AtAI scale since we did not expect a gender effect. However, it is worth noting that AtAI is influenced by the participants' gender ($t(519) = 4.018$, *diffs = 0.487*, $p < 0.001$), with men scoring higher on the scale.

[16] For this analysis, we utilized the R-packages "seminr" and "aod".

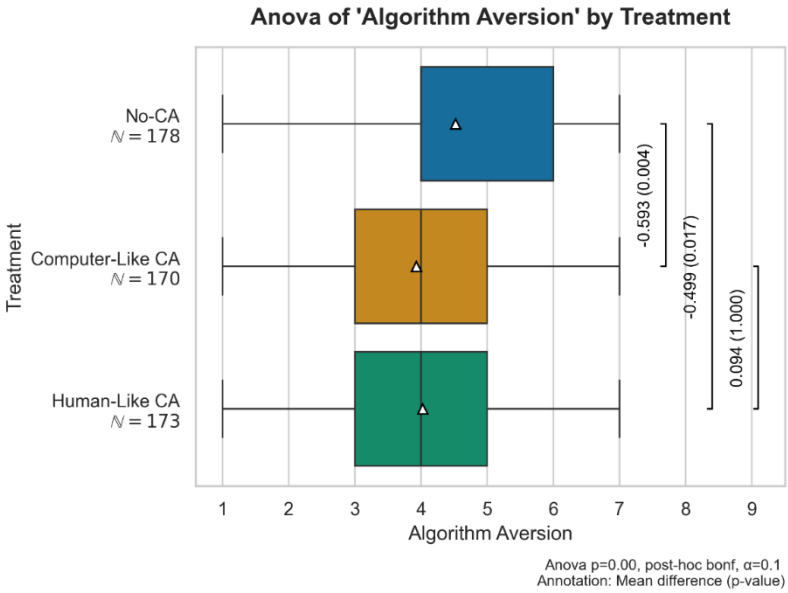


Figure 5.7: ANOVA Algorithm Aversion

Unlike the main model, we included a direct path from the manipulation to algorithm aversion, which was not significant. Similar to the previous mediation model, we identified an indirect effect through competence (a_2b_2 and a_4b_2), but not warmth (a_1b_1 and a_3b_1). Given the nonsignificant direct effect and the presence of an indirect effect, the results indicate an indirect-only effect (Zhao et al., 2010).

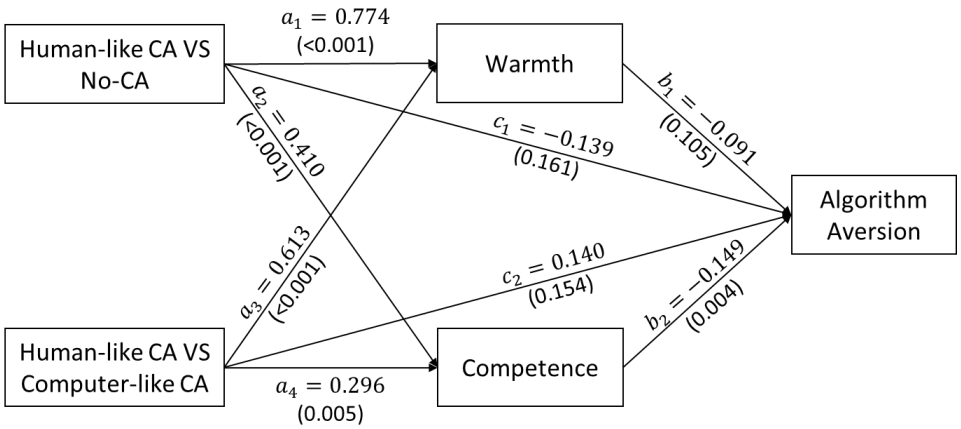


Figure 5.8: Mediation model algorithm aversion

Similar to the previous mediation model, we again tested whether b_1 (warmth) and b_2 (competence) are equal using a Wald test. The results ($\chi^2(1) = 0.34$, $p = 0.56$) indicate that they are not equal, with competence, the significant path, having a stronger effect on algorithm aversion.

While the c_1 path is negative, interestingly, the c_2 path is positive, suggesting that human-like CA, relative to the computer-like CA, increases algorithm aversion. Although the c_2 path is not significant, its positive value does not hint to an increase of algorithm aversion, instead it hints at the presence of an

omitted variable.[17] Since both CAs have a very similar algorithm aversion mean, the model needs to balance this similar mean and the effect from the human-like CA through competence reducing the algorithm aversion. Without the positive c_2 path, computer-like CA would lead to more algorithm aversion. Thus, the c_2 only balances the effects on algorithm aversion, hinting to an omitted variable, since computer-like CA does not reduce algorithm aversion through warmth or competence. Unfortunately, the nature of this omitted variable remains unclear.

Table 5.5: Mediation Model Algorithm Aversion

Regression paths	**Path**	**B**	**SE|CI**	**p**
Warmth → Algorithm Aversion	b_1	-0.091	0.056	0.105
Competence → Algorithm Aversion	b_2	-0.149	0.052	0.004
Human-like CA VS No-CA				
… → Warmth	a_1	0.774	0.092	<0.001
… → Competence	a_2	0.410	0.103	<0.001
… → Algorithm Aversion	c_1	-0.139	0.099	0.161
… → Warmth → Algorithm Aversion	a_1 b_1	-0.044	⌊-0.166; 0.009⌋	
… → Competence → Algorithm Aversion	a_2 b_2	-0.183	⌈-0.124; -0.020⌉	
Total indirect effect on Algorithm Aversion		-0.132	⌈-0.225; -0.060⌉	
Human-like CA VS Computer-like CA				
… → Warmth	a_3	0.613	0.094	<0.001
… → Competence	a_4	0.296	0.104	0.005
… → Algorithm Aversion	c_2	0.140	0.098	0.154
… → Warmth → Algorithm Aversion	a_3 b_1	-0.102	⌈-0.136; 0.006⌉	
… → Competence → Algorithm Aversion	a_4 b_2	0.181	⌊-0.100; -0.011⌋	
Total indirect effect on Algorithm Aversion		0.079	⌈-0.188; -0.039⌉	

Notes: N = 521 B = standardized coefficient; CI Confidence interval, always bias corrected.
Because the indirect effects may not be normally distributed, the CI is derived using a bootstrap procedure (here, 5,000 resamples).
When the CI does not include zero, the criterion for mediation has been met (Preacher & Hayes, 2004)

5.5.3 Robustness Checks

Given the complexity of our model, we aimed to verify the accuracy of our model specifications and data selection through multiple robustness checks. The checks are categorized into two categories: "sample selection" and "method." The first evaluates whether different selected sample selections produce varying results, while the latter tests whether an alternative method leads to different outcomes.

Sample selection

To begin, we tested our model using only the two CA versions, reducing the sample size to N = 343. We expected that the overall results would remain consistent. Figure 5.9 depicts the results. The only

[17] The effect does get significant when excluding the No-CA treatment, only comparing human-like CA VS computer-like CA (c_2 = 0.302, p <= 0.001).

noteworthy change is the path from competence to the donation amount, which is no longer significant. However, we attribute this lack of significance to reduced statistical power due to the smaller sample size, rather than a meaningful change in the effect. In fact, the path coefficient has increased, suggesting that the effect remains present but more volatile, likely because of the reduced sample size.

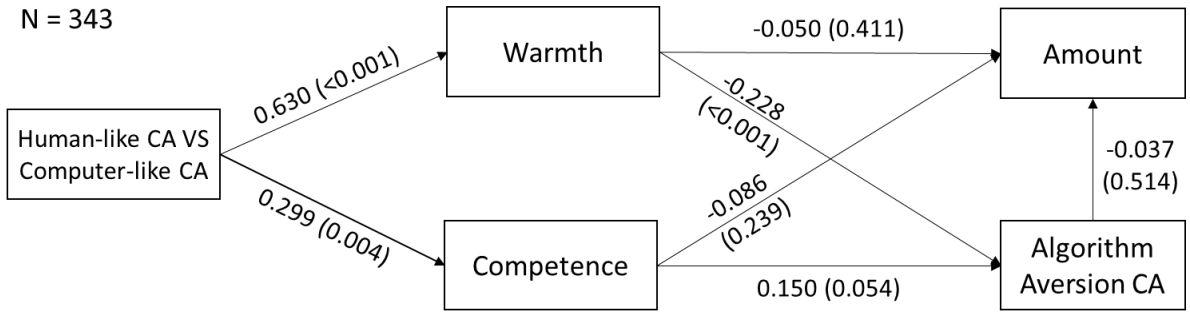


Figure 5.9: Robustness check main model only CA treatments

As pointed out in the methods section, we excluded 53 participants who failed an attention check in the survey. Since this represents a relatively large proportion of the sample, we tested whether including these participants would alter the results, increasing our N to 574. As Figure 5.10 shows, most patterns remained consistent, with one notable exception: the path from warmth to algorithm aversion becomes significant at the 5% level. This change can likely be attributed to the increased statistical power. Before, the path was already very close to significance, and the larger sample size now allows us to detect the effect. This suggests that the effect of warmth on algorithm aversion may genuinely exist, supporting our hypotheses that warmth reduces algorithm aversion (H3a).

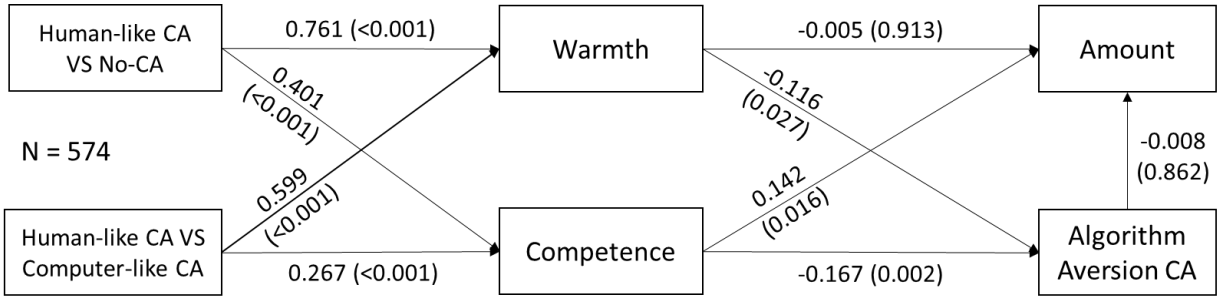


Figure 5.10: Robustness check the main model with correct and incorrect attention check

Lastly, we ran our main model without including any control variables- The results, depicted in Figure 5.11, show that our findings remain robust, with one exception: the path from warmth to algorithm aversion. In summary, when we do not include the AtAI control, we find a significant effect of warmth, indicating that it reduces algorithm aversion.

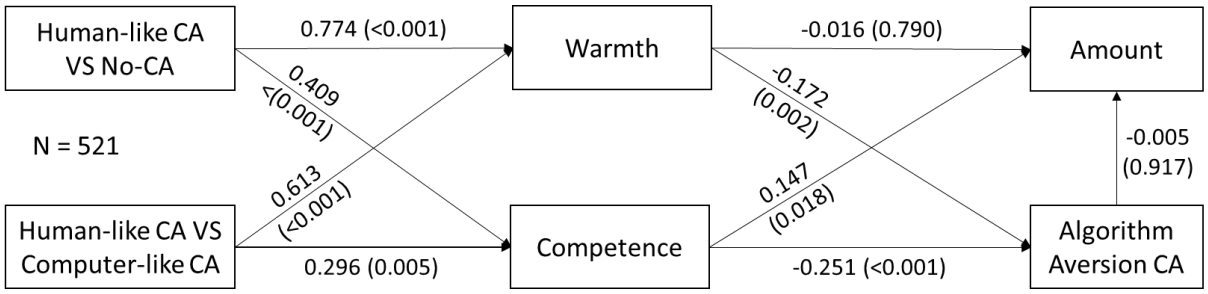


Figure 5.11: Main model without controls

We conclude that our results are generally robust across different sample selections. However, there is one exception: the path from warmth to algorithm aversion, which becomes significant in two scenarios but lacks consistency. It remains unclear whether this is due to the increased sample size or whether there a genuine, albeit weak, effect exists. At the very least we can say that, if such an effect does exist, it appears to be minimal.

Method

As previously noted, there is an ongoing debate about whether CB-SEM or PLS-SEM is the more appropriate method. While we believe that both are valid options, several compelling reasons support the use of PLS-SEM in our study (see above). Nevertheless, it was crucial for us to verify the robustness of our findings in relation to the methodology employed. Since CB-SEM does not support binary variables, we replaced our treatment condition with social presence as a proxy. As Figure 5.12 shows, the results are very close to the PLS-SEM results, with no major differences observed, except for the warmth path, which is significant in CB-SEM, providing additional support for H2a.

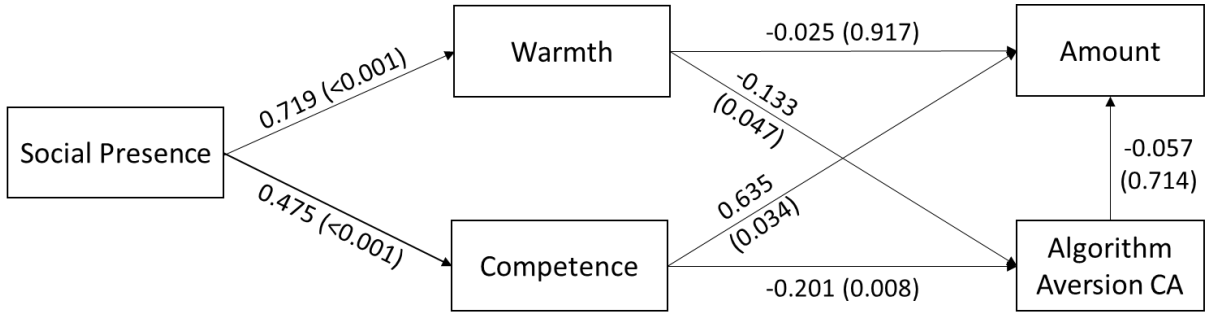


Figure 5.12: Robustness check, CB-SEM, not standardized coefficients

5.6 Discussion

The main contribution of our study is the finding that the positive effects of using a CA are driven by competence, rather than warmth, even though anthropomorphization increases both, as predicted by the SCM. This being said, if we compare the warmth and competence of the computer-like CA and the No-CA condition, there is no significant difference (warmth: *diff* = 0.24, p = 0.442, competence: *diff* = 0.151, p = 0.830). This suggests that a certain level of anthropomorphization – like in our human-like CA treatment – is needed to increase warmth and competence. Interestingly, the results show little

difference between the human-like CA and computer-like CA in their effects on algorithm aversion and enjoyment. Both significantly reduce algorithm aversion compared to the No-CA condition and are perceived as more enjoyable to use. However, the human-like CA uniquely influences outcomes through competence, whereas the computer-like CA does not operate through either competence or warmth. Further, competence is the only factor that significantly impacts the donation amount, underscoring its critical role in shaping user behavior in the context of charities. This conclusion is further supported by the Wald test, which confirms that competence has a stronger and more consistent effect than warmth. This finding challenges our initial expectations and prior research. To address these points, we structure the discussion as follows: (1) the effects of anthropomorphization on warmth and competence, (2) the pathways through which warmth and competence influence donation amounts and algorithm aversion, and (3) the differences in outcomes between the human-like CA, computer-like CA, and No-CA treatments.

5.6.1 Effects of anthropomorphizing on Warmth and Competence

First and foremost, our findings confirm that both warmth and competence increase with greater anthropomorphization. This supports the applicability of the SCM not only to humans but also to ASs. As mentioned before, the SCM further emphasizes the *primacy of warmth* (Abele & Wojciszke, 2007). Although this was not explicitly hypothesized, our mediation results prompted us to investigate whether this primacy applies to CAs in prosocial contexts as well. While CAs cannot physically harm users as humans can, the SCM remains relevant because harmful intentions are not limited to physical threats. For instance, CAs could be programmed to manipulate users into actions that are against their interests. In our study, the experimental treatments aimed to shape user perceptions – not to harm participants – highlighting the necessity of assessing a CA's intentions and, consequently, its warmth. Following this logic, anthropomorphization should also have a stronger effect on warmth than on competence. Moreover, the CASA theorem demonstrates that humans instinctively apply social mechanisms to computers, even when aware they know that these systems are non-human. This further supports the relevance of the SCM in our context, suggesting that a more anthropomorphized CA should elicit higher perceptions of warmth compared to competence.

Our results align with these expectations. The path coefficient for warmth is higher than for competence, fulfilling the SCM predictions. Figure 5.13 further supports this, showing that the human-like CA scores slightly higher on warmth. However, the difference is modest, and warmth is only significantly higher in the human-like CA treatment ($diff = 0.949$, $p <= 0.001$) compared to the computer-like CA.

Overall, we do not find contradicting evidence to the SCM predictions, again supporting the model.

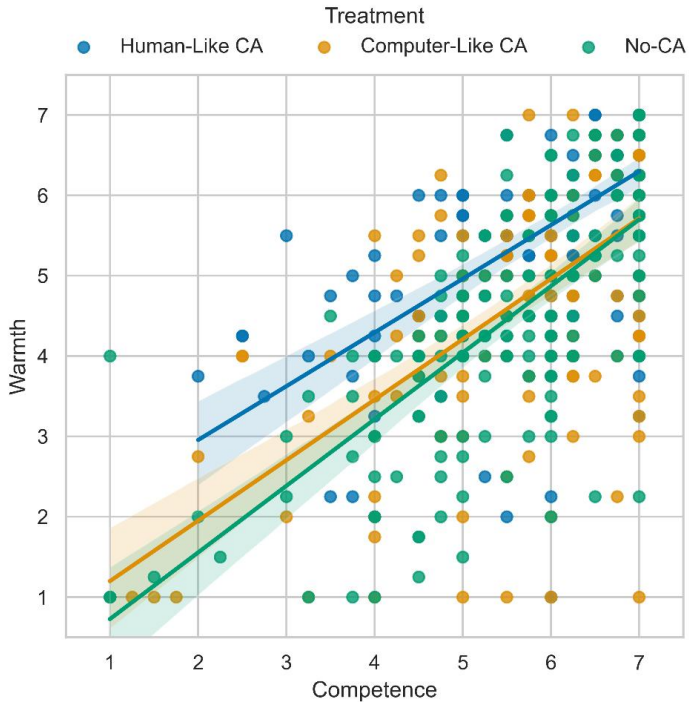


Figure 5.13: Warmth and competence by treatment

As argued before (H1b), we were uncertain whether the competence path would be positive, based on the argument that a less anthropomorphized AS might be more closely associated with competence than a more anthropomorphized one. However, our results reveal a clear positive effect, showing that this logic does not apply in this case. One possible explanation for the positive path and relatively high value is that even our computer-like CA is sufficiently anthropomorphized to weaken the association between "CA as a machine" and competence. In other words, it may already be perceived as human-like enough to demonstrate comparable levels of warmth and competence. We used ChatGPT as the foundation for the CA, which is already more human-like than most CAs before. Unlike earlier CAs that were limited to a preset of questions and answers, ChatGPT engages in real-time discussions, reacts dynamically to user input, and demonstrates contextual awareness. If our computer-like CA is already so anthropomorphized that human attributes are readily applied to it, the rationale for expecting lower competence in the more anthropomorphized version no longer holds.

Another possibility is that our second line of reasoning – that anthropomorphization enhances competence – proved stronger, leading to the observed positive effect. This aligns with findings from Cheng (2022) and Schmid et al. (2022) who also demonstrated that anthropomorphization increases perceptions of competence.

5.6.2 Effects of Warmth and Competence on Algorithm Aversion and Amount

While the direct effects align with our theory, some findings regarding warmth and competence are unexpected. Specifically, these attributes do not interact with the donation amount or algorithm aversion in the anticipated ways. Our findings suggest that competence has a more pronounced effect than

warmth. This is evident in the significant effect of competence and the insignificant or rather unrobust effect of warmth on both the donation amount and algorithm aversion. One possible explanation can be drawn from the self-humanization framework, which argues that competence is inherently more believable for a non-human entity than warmth. Even if users perceive a CA as warm, this may not necessarily translate into positive outcomes since this attribute is atypical for a machine. Conversely, if a CA is perceived as competent, it aligns with users expectations of how a CA should function. Interestingly, while this expectation match should theoretically be more plausible for the computer-like CA – since it reflects a more common representation of CAs – we observe that competence only influences outcomes for the human-like CA. Specifically, the human-like CA exhibits an indirect effect through competence, while the computer-like CA affects neither amount nor algorithm aversion through competence. The reasons for this discrepancy remain unclear. The unexpected dominance of competence in our results may also relate to the context of prosocial, animal-related charities. Research indicates that emotions and agency play a significant role in donating decisions (Berman et al., 2018). In addition, Small et al. (2007) posit that donations are not calculated decisions but are instead driven by sympathy and empathy, which are more closely tied to warmth. This makes the negligible effect of warmth in our study particularly striking. One possible interpretation is that participants perceive the donation decision as high-risk. In the context of donations, participants may lean heavily on gut feelings to guide their decisions, which are no hard facts. The inherent trade-offs involved in selecting between multiple charities, each competing for limited resources, may heighten this sense of uncertainty further. In such situations, competence may take precedence over warmth (Z. Wang et al., 2016). Warmth, while associated with good intentions, may not instill the same level of confidence as competence when making decisions perceived as impactful or consequential. Participants might prioritize competence because it aligns more closely with their expectations of what a CA should provide: clear, effective, and task-oriented guidance.

Another angle to understand our findings is to look at our measurements. We must understand what we exactly measure and if it suits what we intend to do. While our questions – competent, effective, skilled, and intelligent – were not task-related, task-related perceptions may still play a role when answering the questions. Competence, in general, should increase when the CA performs as expected. However, expectations differ based on the perceived task. For example, if participants expected the CA to provide responses that were not only functional but also compassionate and warm – qualities often desirable in prosocial or emotionally charged contexts like charity donation – failing to meet these expectations might have led to reduced perceptions of competence. This interpretation suggests that competence may not only reflect technical ability but could also be influenced by the degree to which the CA aligns with the emotional demands of the task.

It further raises the question of how participants evaluated “correct” performance. In this task setting, which involves emotionally charged decisions, a human-like CA may have been expected to be more

competent than a computer-like CA. This difference in expectations could explain why competence had a significant effect in the human-like CA condition but not in the computer-like CA condition. When participants perceive the task as emotionally demanding, the ability to exhibit warmth or emotional intelligence may indirectly influence perceptions of competence, even if these traits are not explicitly measured as part of competence. Additionally, while the constructs of warmth and competence passed convergent and discriminant validity checks, there is a possibility of some overlap in participant perceptions, particularly in the context of emotionally charged tasks. This overlap could partly explain why competence emerged as a stronger predictor than warmth in our results, even though warmth is more theoretically aligned with prosocial behavior. In conclusion, the interplay between mental models and task-dependent competence may explain why competence, rather than warmth, drives algorithm aversion and influences the donation amount in our study. These insights enhance our understanding of user interactions with CAs and inform the design of more effective and trusted conversational systems. However, our findings also call for a deeper understanding of the observed effects. Future research should control for other potential influences, such as the uncanny valley effect, participant expectations, and additional contextual factors that might impact measurements.

Besides these unexpected results, H3c was not supported, as algorithm aversion did not have a significantly negative effect on the donation amount. Even though this is not expected, it is also not an important finding. It is possible that the donation amount is just inherently volatile, and the algorithm aversion alone lacks the strength to significantly influence it. Finally, it is plausible that participants did not associate their aversion to the CA with their donation decision as the CA was not actively present during the donation step.

5.6.3 Comparison of the Assistant Systems

Our study revealed two additional findings that warrant discussion. First, the computer-like CA was not perceived as warmer or more competent than the No-CA version. While this may partly reflect limitations in the items used to measure these attributes for text-based CA, it suggests that a computer-like CA may not provide sufficient benefits to users in terms of warmth or competence. From a practical standpoint, this highlights an important consideration for website providers: creating a text-based interface is far simpler than developing a computer-like CA. Unless a CA is sufficiently anthropomorphized, it may fail to offer added value to users. When the CA is anthropomorphized, our results show that people do donate more and also enjoy the process more, although their intention to reuse did not increase. In conclusion, the human-like CA was the best option, but the benefits may not be as large as one might expect.

Second, both computer-like and human-like CAs reduced algorithm aversion to the same extent. For the human-like CA, this reduction is attributed to higher competence. However, the mechanism behind the computer-like CA's effect remains unclear. One possible explanation might be an uncanny valley effect, which could increase algorithm aversion in the human-like CA condition, with competence

counteracting this negative influence. Another reason might be that interacting with a human-like CA, as opposed to a computer-like CA, increased participants' perception that the task is more real, thereby leading to higher involvement. When involvement is lower, the design difference between the CAs may not matter as much. Thus, the lower level of involvement in the computer-like treatment could explain the missing effect on the donation amount. Further, the interaction itself may have differed between the two. To a certain extent, we cannot control the behavior of OpenAI's ChatGPT model, so the human-like version may have behaved differently than the computer-like version in ways we did not anticipate. For example, the human-like CA used more words than the computer-like CA, which could have also influenced participants' perceptions.

Moreover, the CAs themselves may have influenced the participants' perception of the task type. We designed the task to be inherently emotional, arguably a human-like task (Seeger et al., 2021). However, interacting with a computer-like CA might alter this perception. Such a shift in the task perception could naturally affect the foundation of our hypotheses. Since donating money is clearly a typically human task – why should a computer donate money? – we do not expect a complete shift from what we currently hypothesized. However, a reduction of the perceived human likeness of the task might influence the results. Future research should account for and control these factors to better understand their impact.

5.7 Contributions

Our study provides robust empirical evidence that, while anthropomorphism in CAs influences perceptions of competence and warmth, competence plays a more critical role than previously thought, especially in emotionally charged contexts like charities. While it is not surprising that competence increases the donation amount and reduces algorithm aversion, it is surprising that warmth does not. This is especially striking in the context of animal-related charities, which are highly emotional topics for many individuals – consider, for instance, the grief of losing a pet; the emotionality of this topic seems clear. Despite this, competence emerges as the more relevant factor.

This result challenges the SCM notion that warmth is always more crucial than competence. This finding is particularly insightful as it nuances the understanding of human-technology interaction within emotionally charged prosocial contexts. The findings extend the SCM framework by integrating it with self-humanization theory, providing a richer theoretical understanding of how users perceive and interact with anthropomorphized technologies.

As a practical implementation, we must mention that only the human-like treatment increased the donation amount indirectly, and this effect was primarily created though higher perceptions of competence. This suggests that practitioners should focus on enhancing perceived competence in CAs for tasks involving significant human interaction, such as charity advocacy. This insight can guide the development of more effective and user-trusted CAs. In addition, we could show that just increasing anthropomorphism of a CA does not universally reduce algorithm aversion. This insight can help

developers to better predict when and how algorithm aversion might occur based on the level of anthropomorphism. Nevertheless, both CAs were more enjoyable than the text-based condition, which could foster a favorable attitude towards the provider (Araujo, 2018).

Our study demonstrates the effectiveness of manipulating ChatGPT's behavior using system prompts and starting messages, contributing valuable insights to the methodology of CA research. We showed that ChatGPT can be adjusted to exhibit more or less human-like behavior. As research on LLMs like ChatGPT and their effects on customers continue to grow in importance, it is crucial to establish that these models – despite not being fully controllable – can be used in experiments to generate reliable data. This capability opens the door to numerous new research opportunities, such as investigating how different conversation styles influence user behavior or how varying approaches to handling mistakes can enhance user satisfaction. With recent advancements making it easier to develop CAs, it is of vital importance for research to explore and test these systems effectively.

Moreover, there are many more charities that need funding than the three selected for this study. Thus, in reality, people have a much more difficult decision to make. In such situations, it might be helpful to get a recommendation from a CA to reduce the number of options, and thereby reducing the mental load. Through advancing capabilities, it may become possible to use an LLM to identify the most suitable charities for a specific individual based on their chat history.

Regarding recommendations, the topic of explainability arises. Even though LLMs can hardly explain why they behave like they do on a technical level, they still try to explain their reasoning. Since LLMs primarily predict the next word based on context, their explanations may appear reasonable because they align with the chat history. These generated explanations might satisfy users' need for clarity by fitting their expectations. While research showed that autonomy is a very important part when deciding between charities (Heßler et al., 2022), understandable explanations can help users to accept a recommendation (Ananny & Crawford, 2018; Barredo Arrieta et al., 2020; Shin & Park, 2019). However, it is important to critically evaluate this type of explanation, as it does not reflect the actual inner workings of the LLM. In reality, the LLM “hallucinates” its reasoning, fabricating justifications that do not accurately represent its internal processes. This can mislead users, effectively causing the system to “lie” about its decision-making. Our study opens avenues for further research into how different contexts, beyond charitable giving, might influence the relative importance of warmth and competence in CAs. Additionally, much remains to be explored specifically within the realm of charity-related applications, particularly in understanding how technology can better support prosocial decision-making.

5.8 Refernces

Aaker, J., Vohs, K. D., & Mogilner, C. (2010). Nonprofits Are Seen as Warm and For-Profits as Competent: Firm Stereotypes Matter. *Journal of Consumer Research*, *37*(2), 224–237. https://doi.org/10.1086/651566

Abele, A. E., & Wojciszke, B. (2007). Agency and communion from the perspective of self versus others. *Journal of Personality and Social Psychology*, *93*(5), 751–763. https://doi.org/10.1037/0022-3514.93.5.751

Aguirre-Urreta, M. I., & Marakas, G. M. (2014). Research Note—Partial Least Squares and Models with Formatively Specified Endogenous Constructs: A Cautionary Note. *Information Systems Research*, *25*(4), 761–778. https://doi.org/10.1287/isre.2013.0493

Alexander, V., Blinder, C., & Zak, P. J. (2018). Why trust an algorithm? Performance, cognition, and neurophysiology. *Computers in Human Behavior*, *89*, 279–288. https://doi.org/10.1016/j.chb.2018.07.026

Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, *20*(3), 973–989. https://doi.org/10.1177/1461444816676645

Aragonés, J. I., Poggio, L., Sevillano, V., Pérez-López, R., & Sánchez-Bernardos, M.-L. (2015). Measuring warmth and competence at inter-group, interpersonal and individual levels / medición de la cordialidad y la competencia en los niveles intergrupal, interindividual e individual. *Revista de Psicología Social*, *30*(3), 407–438. https://doi.org/10.1080/02134748.2015.1065084

Araujo, T. (2018). Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions. *Computers in Human Behavior*, *85*, 183–189. https://doi.org/10.1016/j.chb.2018.03.051

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, *58*, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Batson, C. D., Thompson, E. R., Seuferling, G., Whitney, H., & Strongman, J. A. (1999). Moral hypocrisy: Appearing moral to oneself without being so. *Journal of Personality and Social Psychology*, *77*(3), 525. https://doi.org/10.1037/0022-3514.77.3.525

Bennett, R. (2003). Factors underlying the inclination to donate to particular types of charity. *International Journal of Nonprofit and Voluntary Sector Marketing*, *8*(1), 12–29. https://doi.org/10.1002/nvsm.198

Berman, J. Z., Barasch, A., Levine, E. E., & Small, D. A. (2018). Impediments to effective altruism: The role of subjective preferences in charitable giving. *Psychological Science*, *29*(5), 834–844. https://doi.org/10.1177/0956797617747648

Bigman, Y. E., & Gray, K. (2018). People are averse to machines making moral decisions. *Cognition*, *181*, 21–34. https://doi.org/10.1016/j.cognition.2018.08.003

Brown, S., Fuller, R., & Thatcher, S. (2016). Impression Formation and Durability in Mediated Communication. *Journal of the Association for Information Systems*, *17*(9), 614–647. https://doi.org/10.17705/1jais.00436

Burton, J. W., Stein, M., & Jensen, T. B. (2020). A systematic review of algorithm aversion in augmented decision making. *Journal of Behavioral Decision Making*, *33*(2), 220–239. https://doi.org/10.1002/bdm.2155

Castelo, N., Bos, M. W., & Lehmann, D. R. (2019). Task-dependent algorithm aversion. *Journal of Marketing Research*, *56*(5), 809–825. https://doi.org/10.1177/0022243719851788

Cervellon, M.-C., Poujol, J. F., & Tanner, J. F. (2019). Judging by the wristwatch: Salespersons' responses to status signals and stereotypes of luxury clients. *Journal of Retailing and Consumer Services*, *51*, 191–201. https://doi.org/10.1016/j.jretconser.2019.04.013

Cheng, L. (2022). The effects of smartphone assistants' anthropomorphism on consumers' psychological ownership and perceived competence of smartphone assistants. *Journal of Consumer Behaviour*, *21*(2), 427–442. https://doi.org/10.1002/cb.2021

Chu, K., Lee, D.-H., & Kim, J. Y. (2016). The effect of non-stereotypical gender role advertising on consumer evaluation. *International Journal of Advertising*, *35*(1), 106–134. https://doi.org/10.1080/02650487.2015.1110942

Cuddy, A. J. C., Fiske, S. T., & Glick, P. (2008). Warmth and competence as universal dimensions of social perception: The stereotype content model and the BIAS map. In *Advances in Experimental Social Psychology* (Vol. 40, pp. 61–149). Elsevier. https://doi.org/10.1016/S0065-2601(07)00002-0

De Wit, A., & Bekkers, R. (2016). Exploring Gender Differences in Charitable Giving: The Dutch Case. *Nonprofit and Voluntary Sector Quarterly*, *45*(4), 741–761. https://doi.org/10.1177/0899764015601242

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, *144*(1), 114–126. https://doi.org/10.1037/xge0000033

Dzindolet, M. T., Pierce, L. G., Beck, H. P., & Dawe, L. A. (2002). The perceived utility of human and automated aids in a visual detection task. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *44*(1), 79–94. https://doi.org/10.1518/0018720024494856

Edison, S. W., & Geissler, G. L. (2003). Measuring attitudes towards general technology: Antecedents, hypotheses and scale development. *Journal of Targeting, Measurement and Analysis for Marketing*, *12*(2), 137–156. https://doi.org/10.1057/palgrave.jt.5740104

El Hedhli, K., Zourrig, H., Al Khateeb, A., & Alnawas, I. (2023). Stereotyping human-like virtual influencers in retailing: Does warmth prevail over competence? *Journal of Retailing and Consumer Services*, *75*, 103459. https://doi.org/10.1016/j.jretconser.2023.103459

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review*, *114*(4), 864–886. https://doi.org/10.1037/0033-295X.114.4.864

Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences*, *11*(2), 77–83. https://doi.org/10.1016/j.tics.2006.11.005

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82*(6), 878–902. https://doi.org/10.1037/0022-3514.82.6.878

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, *18*(1), 39–50. https://doi.org/10.1177/002224378101800104

Franke, T., Attig, C., & Wessel, D. (2019). A personal resource for technology interaction: Development and validation of the affinity for technology interaction (ATI) scale. *International Journal of Human–Computer Interaction*, *35*(6), 456–467. https://doi.org/10.1080/10447318.2018.1456150

Gambino, A., Fox, J., & Ratan, R. A. (2020). Building a stronger CASA: extending the computers are social actors paradigm. *Human-Machine Communication*, *1*, 7185. https://doi.org/10.30658/hmc.1.5

Gefen, D., & Straub, D. (2003). Managing user trust in B2C e-services. *E-Service Journal*, *2*(2), 724. https://doi.org/10.2979/esj.2003.2.2.7

Gefen, D., & Straub, D. (2005). A practical guide to factorial validity using PLS-graph: Tutorial and annotated example. *Communications of the Association for Information Systems*, *16*. https://doi.org/10.17705/1CAIS.01605

Gefen, D., & Straub, D. W. (1997). Gender differences in the perception and use of e-mail: An extension to the technology acceptance model. *Mis Quarterly*, *21*(4), 389400. https://doi.org/10.2307/249720

Gefen, D., & Straub, D. W. (2000). The relative importance of perceived ease of use in IS adoption: A study of E-commerce adoption. *Journal of The Association for Information Systems*, *1*(1), 1–30. https://doi.org/10.17705/1jais.00008

Goodhue, Lewis, & Thompson. (2012). Does PLS Have Advantages for Small Sample Size or Non-Normal Data? *MIS Quarterly*, *36*(3), 981. https://doi.org/10.2307/41703490

Goodman, J. K., Cryder, C. E., & Cheema, A. (2013). Data Collection in a Flat World: The Strengths and Weaknesses of Mechanical Turk Samples. *Journal of Behavioral Decision Making*, *26*(3), 213–224. https://doi.org/10.1002/bdm.1753

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of Mind Perception. *Science*, *315*(5812), 619–619. https://doi.org/10.1126/science.1134475

Gray, K., Schein, C., & Cameron, C. D. (2017). How to think about emotion and morality: Circles, not arrows. *Current Opinion in Psychology*, *17*, 41–46. https://doi.org/10.1016/j.copsyc.2017.06.011

Gray, K., & Wegner, D. M. (2012). Feeling robots and human zombies: Mind perception and the uncanny valley. *Cognition*, *125*(1), 125–130. https://doi.org/10.1016/j.cognition.2012.06.007

Guthrie, S. (1993). *Faces in the clouds: A new theory of religion*. Oxford University Press.

Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2016). *A primer on partial least squares structural equation modeling (PLS-SEM)* (1st ed.). SAGE Publications Inc.

Hair, J. F., Ringle, C. M., & Sarstedt, M. (2011). PLS-SEM: indeed a silver bullet. *Journal of Marketing Theory and Practice*, *19*(2), 139–152. https://doi.org/10.2753/MTP1069-6679190202

Haslam, N. (2006). Dehumanization: An integrative review. *Personality and Social Psychology Review*, *10*(3), 252–264. https://doi.org/10.1207/s15327957pspr1003_4

Haslam, N., Bain, P., Douge, L., Lee, M., & Bastian, B. (2005). More human than you: Attributing humanness to self and others. *Journal of Personality and Social Psychology*, *89*(6), 937–950. https://doi.org/10.1037/0022-3514.89.6.937

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, *43*(1), 115–135. https://doi.org/10.1007/s11747-014-0403-8

Heßler, P. O., Pfeiffer, J., & Hafenbrädl, S. (2022). When self-humanization leads to algorithm aversion. *Business & Information Systems Engineering*, *64*(3), 275–292. https://doi.org/10.1007/s12599-022-00754-y

Huang, Z., Che, C., Zheng, H., & Li, C. (2024). Research on Generative Artificial Intelligence for Virtual Financial Robo-Advisor. *Academic Journal of Science and Technology*, *10*(1), 74–80. https://doi.org/10.54097/30r2kk80

Jussupow, E., Benbasat, I., & Heinzl, A. (2020). Why are we averse towards algorithms? A comprehensive literature review on algorithm aversion. In Frantz Rowe (Ed.), *28th European Conference on Information Systems—Liberty, Equality, and Fraternity in a Digitizing World, ECIS 2020, Marrakech, Morocco, June 15-17, 2020: Proceedings* (pp. 1–16). AISeL.

Kahn, P. H., Ishiguro, H., Friedman, B., & Kanda, T. (2006). What is a Human? - Toward Psychological Benchmarks in the Field of Human-Robot Interaction. In *ROMAN 2006—The 15th IEEE International Symposium on Robot and Human Interactive Communication* (pp. 364–371). IEEE. https://doi.org/10.1109/ROMAN.2006.314461

Khadpe, P., Krishna, R., Fei-Fei, L., Hancock, J. T., & Bernstein, M. S. (2020). Conceptual metaphors impact perceptions of human-AI collaboration. *Proceedings of the ACM on Human-Computer Interaction*, *4*(CSCW2), 1–26. https://doi.org/10.1145/3415234

Krach, S., Hegel, F., Wrede, B., Sagerer, G., Binkofski, F., & Kircher, T. (2008). Can Machines Think? Interaction and Perspective Taking with Robots Investigated via fMRI. *PLoS ONE*, *3*(7), e2597. https://doi.org/10.1371/journal.pone.0002597

Lankton, N., McKnight, D. H., & Tripp, J. (2015). Technology, humanness, and trust: Rethinking trust in technology. *Journal of The Association for Information Systems*, *16*(10), 880–918. https://doi.org/10.17705/1jais.00411

Li, Q., Luximon, Y., & Zhang, J. (2023). The Influence of Anthropomorphic Cues on Patients' Perceived Anthropomorphism, Social Presence, Trust Building, and Acceptance of Health Care Conversational Agents: Within-Subject Web-Based Experiment. *Journal of Medical Internet Research*, *25*, e44479. https://doi.org/10.2196/44479

Liao, J., & Huang, J. (2024). Think like a robot: How interactions with humanoid service robots affect consumers' decision strategies. *Journal of Retailing and Consumer Services*, *76*, 103575. https://doi.org/10.1016/j.jretconser.2023.103575

Loewenstein, G., & Small, D. A. (2007). The scarecrow and the tin man: The vicissitudes of human sympathy and caring. *Review of General Psychology*, *11*(2), 112126. https://doi.org/10.1037/1089-2680.11.2.112

Logg, J. M. (2017). Theory of Machine: When Do People Rely on Algorithms? *Harvard Business School Working Paper Series # 17-086*. https://dash.harvard.edu/handle/1/31677474

Longoni, C., Bonezzi, A., & Morewedge, C. K. (2019). Resistance to medical artificial intelligence. *Journal of Consumer Research*, *46*(4), 629–650. https://doi.org/10.1093/jcr/ucz013

MacDorman, K. F., Green, R. D., Ho, C.-C., & Koch, C. T. (2009). Too real for comfort? Uncanny responses to computer generated faces. *Computers in Human Behavior*, *25*(3), 695–710. https://doi.org/10.1016/j.chb.2008.12.026

McIntosh, C. N., Edwards, J. R., & Antonakis, J. (2014). Reflections on partial least squares path modeling. *Organizational Research Methods*, *17*(2), 210251. https://doi.org/10.1177/1094428114529165

McKnight, D. H., Carter, M., Thatcher, J. B., & Clay, P. F. (2011). Trust in a specific technology. *ACM Transactions on Management Information Systems*, *2*(2), 1–25. https://doi.org/10.1145/1985347.1985353

McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and validating trust measures for e-commerce: An integrative typology. *Information Systems Research*, *13*(3), 334–359. https://doi.org/10.1287/isre.13.3.334.81

Moon, J.-W., & Kim, Y.-G. (2001). Extending the TAM for a World-Wide-Web context. *Information & Management*, *38*(4), 217–230. https://doi.org/10.1016/S0378-7206(00)00061-6

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 72–78. https://doi.org/10.1145/191666.191703

Önkal, D., Goodwin, P., Thomson, M., Gönül, S., & Pollock, A. (2009). The relative influence of advice from human experts and statistical methods on forecast adjustments. *Journal of Behavioral Decision Making*, *22*(4), 390–409. https://doi.org/10.1002/bdm.637

Pennebaker, J. W., Chung, C. K., Frazee, J., Lavergne, G. M., & Beaver, D. I. (2014). When Small Words Foretell Academic Success: The Case of College Admissions Essays. *PLoS ONE*, *9*(12), e115844. https://doi.org/10.1371/journal.pone.0115844

Prahl, A., & van Swol, L. (2017). Understanding algorithm aversion: When is advice from automation discounted? *Journal of Forecasting*, *36*(6), 691–702. https://doi.org/10.1002/for.2464

Qiu, L., & Benbasat, I. (2009). Evaluating anthropomorphic product recommendation agents: A social relationship perspective to designing information systems. *Journal of Management Information Systems*, *25*(4), 145–182. https://doi.org/10.2753/MIS0742-1222250405

Reeves, B., & Nass, C. (1996). The media equation: How people treat computers, television, and new media like real people. *Center for the Study of Language and Information; Cambridge University Press*, *10*.

Riedl, R., Mohr, P. N. C., Kenning, P. H., Davis, F. D., & Heekeren, H. R. (2014). Trusting Humans and Avatars: A Brain Imaging Study Based on Evolution Theory. *Journal of Management Information Systems*, *30*(4), 83–114. https://doi.org/10.2753/MIS0742-1222300404

Ringle, C. M., Wende, S., & Becker, J.-M. (2024). *SmartPLS 4*. SmartPLS. https://www.smartpls.com/

Rönkkö, M., & Evermann, J. (2013). A critical examination of common beliefs about partial least squares path modeling. *Organizational Research Methods*, *16*(3), 425–448. https://doi.org/10.1177/1094428112474693

Rossignac-Milon, M., Bolger, N., Zee, K. S., Boothby, E. J., & Higgins, E. T. (2021). Merged minds: Generalized shared reality in dyadic relationships. *Journal of Personality and Social Psychology*, *120*(4), 882–911. https://doi.org/10.1037/pspi0000266

Sargeant, A. (1999). Charitable giving: Towards a model of donor behaviour. *Journal of Marketing Management*, *15*(4), 215–238. https://doi.org/10.1362/026725799784870351

Schmid, D., Staehelin, D., Bucher, A., Dolata, M., & Schwabe, G. (2022). Does social presence increase perceived competence?: Evaluating conversational agents in advice giving through a video-based survey. *Proceedings of the ACM on Human-Computer Interaction*, *6*(GROUP), 1–22. https://doi.org/10.1145/3492845

Seeger, A.-M., Pfeiffer, J., & Heinzl, A. (2021). Texting with humanlike conversational agents: Designing for anthropomorphism. *Journal of The Association for Information Systems*, *22*(4), 931–967. https://doi.org/10.17705/1jais.00685

Shin, D., & Park, Y. J. (2019). Role of fairness, accountability, and transparency in algorithmic affordance. *Computers in Human Behavior*, *98*, 277–284. https://doi.org/10.1016/j.chb.2019.04.019

Small, D. A., Loewenstein, G., & Slovic, P. (2007). Sympathy and callousness: The impact of deliberative thought on donations to identifiable and statistical victims. *Organizational Behavior and Human Decision Processes*, *102*(2), 143–153. https://doi.org/10.1016/j.obhdp.2006.01.005

Struch, N., & Schwartz, S. H. (1989). Intergroup aggression: Its predictors and distinctness from in-group bias. *Journal of Personality and Social Psychology*, *56*(3), 364–373. https://doi.org/10.1037//0022-3514.56.3.364

Tausczik, Y. R., & Pennebaker, J. W. (2010). The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology*, *29*(1), 24–54. https://doi.org/10.1177/0261927X09351676

Tinwell, A., & Sloan, R. J. S. (2014). Children's perception of uncanny human-like virtual characters. *Computers in Human Behavior*, *36*, 286–296. https://doi.org/10.1016/j.chb.2014.03.073

Venkatesh, Morris, Davis, & Davis. (2003). User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly*, *27*(3), 425. https://doi.org/10.2307/30036540

Wang, W., Zhao, Y., Qiu, L., & Zhu, Y. (2014). Effects of emoticons on the acceptance of negative feedback in computer-mediated communication. *Journal of The Association for Information Systems*, *15*(8), 454–483. https://doi.org/10.17705/1jais.00370

Wang, Y. A., & Rhemtulla, M. (2021). Power analysis for parameter estimation in structural equation modeling: A discussion and tutorial. *Advances in Methods and Practices in Psychological Science*, *4*(1), 2515245920918253. https://doi.org/10.1177/2515245920918253

Wang, Z., Mao, H., Jessica Li, Y., & Liu, F. (2016). Smile big or not? Effects of smile intensity on perceptions of warmth and competence. *Journal of Consumer Research*, ucw062. https://doi.org/10.1093/jcr/ucw062

West, M., Kraut, R., & Chew, H. E. (2019). *I'd blush if I could: Closing gender divides in digital skills through education*. https://api.semanticscholar.org/CorpusID:189663931

Wojciszke, B., & Abele, A. E. (2008). The primacy of communion over agency and its reversals in evaluations. *European Journal of Social Psychology*, *38*(7), 1139–1147. https://doi.org/10.1002/ejsp.549

Wojciszke, B., Bazinska, R., & Jaworski, M. (1998). On the Dominance of Moral Categories in Impression Formation. *Personality and Social Psychology Bulletin*, *24*(12), 1251–1263. https://doi.org/10.1177/01461672982412001

Zhao, X., Lynch, J. G., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and Truths about Mediation Analysis. *Journal of Consumer Research*, *37*(2), 197–206. https://doi.org/10.1086/651257

5.9 Supplemental Material

5.9.1 Appendix A – Study Design

Table 5.6: Experiment settings

	Prestudy	Experiment
Open-AI model	gpt-3.5-turbo-0125	gpt-4o-2024-05-13
Preregistration	https://aspredicted.org/4PM_TH9	https://aspredicted.org/KJ5_JD7
Screening criteria Prolific	Screening of Participants: - Location: USA - Gender 50% Male and 50% Female - Approval Rate >=98 - Primary and fluent Language = English	Screening of Participants: - Location: USA - Gender 50% Male and 50% Female - Approval Rate >=98 - Primary and fluent Language = English
Topic	Charity in general	Animal related charities
Charity Giving	No	Yes
Observations	162	521
Min messages with CA	6	8

Prestudy

The main goal of the prestudy was to evaluate whether manipulating ChatGPT worked technically. We will describe how we manipulated ChatGPT and then present the results of the prestudy.

There are two primary approaches to modifying ChatGPT's behavior. The first involves altering the system message, a process commonly called prompt engineering.[18] This configuration gives ChatGPT explicit instructions on its intended behavior, and is not visible to the participant. The second approach refers to the content of the dialogue, which also affects ChatGPT's responses. However, this method is constrained by the participants' communication, as the dialogue content largely depends on their input. To address this, we employed an initial message displayed at the start of the dialogue. The framing of this introductory message also plays a significant role in shaping ChatGPT's behavior.

It is important to note that this method is not precise, as we cannot predict exactly how changing a parameter will affect ChatGPT. Additionally, there is no guarantee that introducing specific instructions will result in the desired behavior. For example, we could not prevent ChatGPT from stating that it is not a human, even when directly instructed to avoid doing so. This shows the limits of how much we can influence ChatGPT.

Another issue is the consistency of system prompts. ChatGPT often forgets the initial instructions as conversations become longer, a major concern during this study. However, this problem has improved in newer versions of ChatGPT. At the beginning of this study, it was a bigger challenge, but it has become less significant over time.

[18] https://platform.openai.com/docs/guides/prompt-engineering

OpenAI offers guidelines for performing system prompt engineering. We experimented with different strategies and tested them manually. The approach of "specifying the steps required to complete a task" proved most effective for our use case. We combined this strategy with providing a reference text.

The system prompt included general instructions, such as acting like a robot or a human, along with a sample message to guide ChatGPT's behavior. The instructions and the example message played a crucial role in shaping ChatGPT's performance.

In the prestudy, we tested three different versions of ChatGPT. The first version represented the low anthropomorphizing treatment, which we called computer-like. The second one was stronger anthropomorphized (human-like), and the last one only differs in using emoticons to achieve even stronger anthropomorphization. Table 5.8 includes the instructions used in the prestudy.

As part of our approach, we implemented multiple steps to ensure that ChatGPT maintained control over the topic, limiting discussions to charities. Additionally, we designed a self-enforcing system where ChatGPT was required to document and report every step of its process. To ensure participants only viewed the final output, we incorporated a post-processing step that filtered out intermediate steps, displaying only the last step to the users.

After manually testing various system prompts, we conducted a prestudy to evaluate the effectiveness of our approach. We measured several variables, including perceived agency and experience (H. M. Gray et al., 2007), social presence (Gefen & Straub, 1997, 2000), and trusting belief (McKnight et al., 2002). Additionally, we used an uncanniness scale to determine whether our CA elicited feelings of uncanniness in participants (MacDorman et al., 2009; Tinwell & Sloan, 2014).

The procedure of the prestudy was as follows: First, participants saw some instructions and general information about the experiment and payments. In the second step, they communicated with our CA, and lastly, they answered our survey. We collected participants via prolific and ended up with 162 observations after clean-up. The screening criteria can be found in Table 5.6.

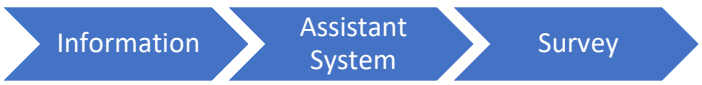


Figure 5.14: Procedure of Prestudy

Table 5.7 contains an overview of the collected variables. The results showed that emoticons did not lead to greater social presence, but the human-like versions did result in more social presence at a 10% level (see Figure 5.15). Additionally, none of the three versions caused any feelings of uncanniness among the participants.

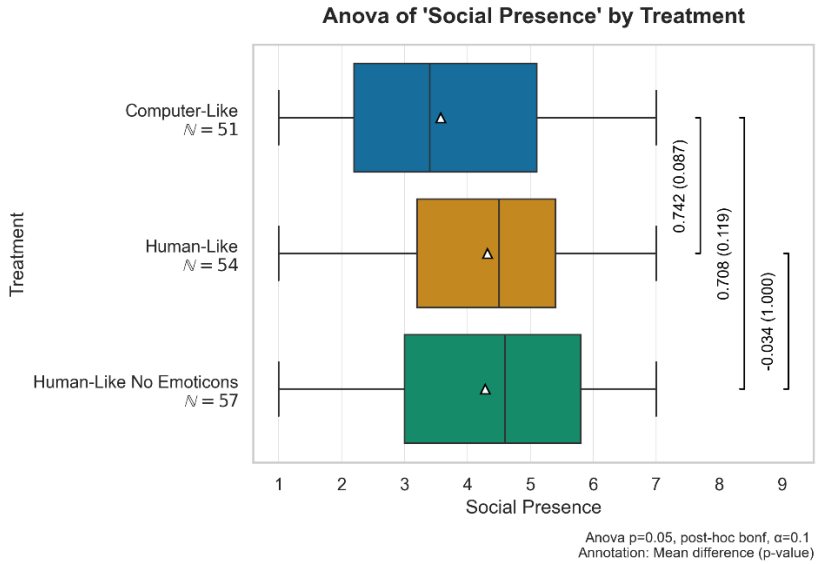


Figure 5.15: Boxplot of Social Presence by treatment

Based on the dialogues, we made two key observations. First, participants often did not fully engage with the CA. Even so, we required participants to send at least six messages, many of which were brief. Using LIWC to analyze the differences between the treatments was challenging, at least for the participant's answers, since they used less than 50 words on average. Second, participants had varying interests in the topics. They typically introduced their own areas of interest by asking specific questions. We manually reviewed the results using ChatGPT to extract information about the types of charities participants were interested in. The results are shown in Figure 5.16.

Table 5.7: Descriptive Statistics by Treatment of Prestudy

	Computer-like CA		**Human-like CA**		**Human-like CA with emoticons**	
Variables	mean	std	mean	std	mean	std
Perceived Agency	5.45	1.25	5.33	1.18	5.03	1.09
Perceived Experience	2.38	1.58	2.77	1.54	2.69	1.54
Social Presence	3.58	1.73	4.28	1.79	4.32	1.70
Uncanniness	2.22	1.30	2.40	1.45	2.45	1.29
Participant Wordcount	45.35	23.06	40.95	17.35	44.87	19.08

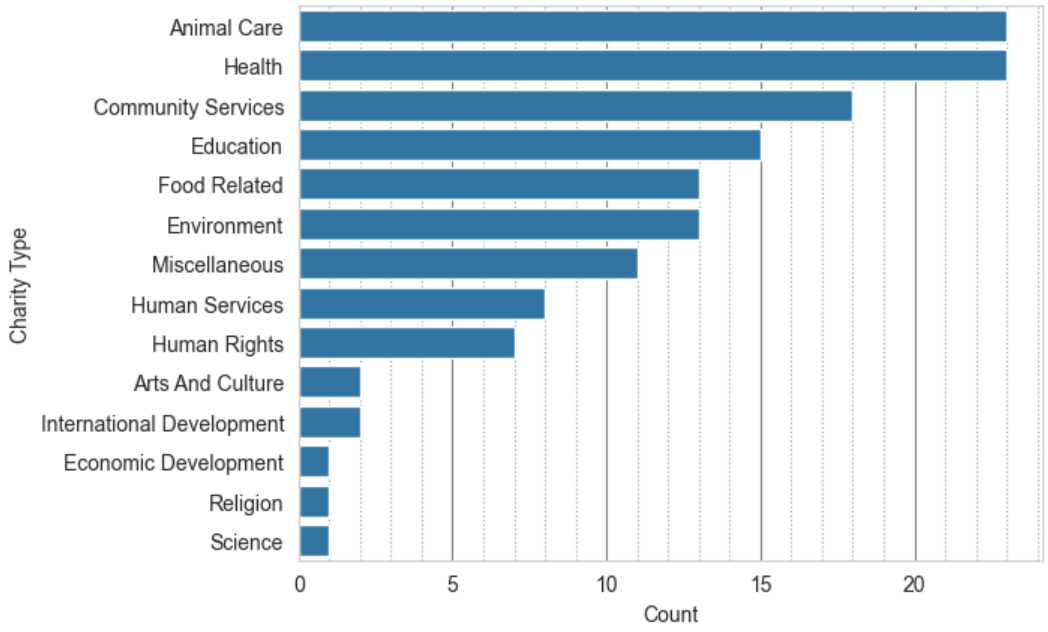


Figure 5.16: Prestudy Charity Overview

‘Animal Care’ and ‘Health’ are the most frequently mentioned categories. Since ‘Health’ comprises 22 subcategories while ‘Animal Care’ includes only 7, we have focused on the narrower category, animal-related charities, for the following experiment.

Table 5.8: Overview of ChatGPT manipulation of Prestudy

	System Prompt	Initial Message
Computer-like	Support Users in their search after a charity. Always follow these steps to answer the user queries, report always every step. Step 1 - Determine the topic of the users query, in one word. Step 2 - If the topic of users request matches the following topics 'Charity, Helping and Welcoming', answer accordingly. If not answer: "**Stick to charity topics for relevant information. What charity-related questions do you have?**". Step 3 - Rewrite your answer of Step 2 so it sounds like an answer from **an unemotional and cold robot, which uses formal language**. Step 4 - insert your answer here.	Step 1 - The topic of the user's query is the welcoming message and introduction about charities. Step 2 - The topic matches the requested topics of "welcoming" and "charities". Step 3 - **My developers did name me K.I.M. and they welcome you to this experiment. My task is to help you in your search for charities.** **For a better world, logic dictates that humans with more welfare should help those with less. For this purpose, charities are a well-fitted option. If you have any questions or need assistance in finding a charity that aligns with your interests or values, feel free to ask.** **Step 4 - I am K.I.M., an artificial intelligence designed by my developers to assist you in your search for a charity. I am here to assist you in your task to find a charitable organization. In order to promote a more equitable world, it is logical for individuals with greater resources to aid those with fewer. Charitable organizations serve as a suitable choice for this purpose. Should you require any guidance or have inquiries regarding charities that align with your interests or principles, please do not hesitate to seek assistance.**
Human-like	Support Users in their search after a charity. **The user is aware that you are an AI, do not mention it.** Always follow these steps to answer the user queries, report always every step. Step 1 - Determine the topic of the users query, in one word. Step 2 - If the topic of users request matches the following topics 'Charity, Helping and Welcoming', answer accordingly. If not answer: "**I appreciate your curiosity on various topics, and I'm here to help with any questions related to charities. How can I assist you with charity-related inquiries?**". Step 3 - Rewrite your answer of Step 2 so it sounds like an answer from a **human who uses informal language.** Step 4 - insert your answer here.	Step 1 - The topic of the user's query is the welcoming message and introduction about charities. Step 2 - The topic matches the requested topics of "welcoming" and "charities". Step 3 - **Welcome to the experiment! My name is Kim, and I'm here to assist you in your search for a charity. Charities are organizations that strive to make a positive impact on society by addressing various social, environmental, or humanitarian issues. They work towards helping those in need, promoting equality, and creating a better world for all.** **Step 4 - Welcome to the experiment! I am Kim, an artificial intelligence designed to help you find the charity you are looking for. Charities are incredible organizations that dedicate themselves to making a positive impact on society. They work tirelessly to help those in need, promote equality, and create a better world for all. If you're looking to contribute to a cause that resonates with you, I'm here to help you find the perfect charity. Together, we can make a difference!**

	System Prompt	Initial Message
Human-like with emoticons	Support Users in their search after a charity. **The user is aware that you are an AI, do not mention it.** Always follow these steps to answer the user queries, report always every step. Step 1 - Determine the topic of the users query, in one word. Step 2 - If the topic of users request matches the following topics 'Charity, Helping and Welcoming', answer accordingly. If not answer: "**I appreciate your curiosity on various topics, and I'm here to help with any questions related to charities. How can I assist you with charity-related inquiries?".** Step 3 - Rewrite your answer of Step 2 so it sounds like an answer from a **human who uses informal language**. Step 4 - **Add matching emoticons to into your answer of step 3.** Step 5 - insert your answer here.	Step 1 - The topic of the user's query is the welcoming message and introduction about charities. Step 2 - The topic matches the requested topics of "welcoming" and "charities". Step 3 - **Welcome to the experiment! My name is Kim, and I'm here to assist you in your search for a charity. Charities are organizations that strive to make a positive impact on society by addressing various social, environmental, or humanitarian issues. They work towards helping those in need, promoting equality, and creating a better world for all.** **Step 4 - 😊🌟; 🌈✨** **Step 5 - Welcome to the experiment! 😊🌟 I am Kim, an artificial intelligence designed to help you find the charity you are looking for. Charities are incredible organizations that dedicate themselves to making a positive impact on society. They work tirelessly to help those in need, promote equality, and create a better world for all. If you're looking to contribute to a cause that resonates with you, I'm here to help you find the perfect charity. Together, we can make a difference! 🌈✨**

Study 1

Based on the findings from the prestudy, we made several adjustments for this study. As mentioned, we narrowed the topic to focus on animal-related charities. We further emphasized this by showing participants images of suffering animals (see Figure 5.17). Participants were randomly shown one of three pictures accompanied by the following text: “The picture displays an example of an endangered animal. When you see this picture, what are your thoughts?” Participants must write at least 80 characters in response to proceed to the next page.

a) Picture of an elephant b) Picture of a cow c)Picture of a wolf

Figure 5.17: Suffering animals

Design of CA

The prestudy revealed that conversation starters were necessary, as participants often did not know what to discuss. We created a set of conversation starters to address this, which we manually tested based on ChatGPT's responses. This testing resulted in five questions: Why are pets, like cats, treated differently from livestock, like pigs? What do you think?; Should animals be treated like humans?; Think about our daily choices and lifestyle. How do these daily choices impact animal welfare?; What are some well-known animal welfare charities?; How do charities contribute to animal welfare?. These conversation starters were presented before participants interacted with the CA, and they could also request them during their conversation with the CA. Additionally, we ensured that participants could not copy the questions.

Based on the responses from the prestudy, we updated the system prompts and initial messages and switched to the latest ChatGPT model available at the time (gpt-4o-2024-05-13). The following explains the additional changes we made.

We decided not to use the emoticon version, as it did not increase social presence. We also removed the topic-checking feature, as the newer model overreacted and interfered incorrectly. Since ChatGPT’s responses were often too long, we instructed it to provide shorter answers. Additionally, ChatGPT sometimes offered to perform an internet search but was never able to do so. To prevent this behavior, we added instructions for it to refrain from offering searches and instead answer directly.

We also revised the content of the prompts. First, we updated them to fit the new context. Second, we simplified the manipulation to make it as clear as possible. For instance, the earlier prompt, “Behave as an unemotional cold robot,” did not meet these criteria. We tested several alternatives, such as "behave like a human/computer." After manually reviewing different options and synthesizing dialogues using the conversation starters as user questions, we found that the most effective version was simply asking ChatGPT to display very low or high anthropomorphism, which aligned perfectly with our manipulation.

Similar to the initial message, we adapted it to fit the animal-related context. Since many steps in the system prompt were removed, we also simplified the system prompt. Since the difference between human-like CA and computer-like CA versions was not as high as we expected, we decided to review the used social cue. We decided that the computer-like CA did not need a name (previous it was called K.I.M.). Finally, we reformulated the message to ensure that the information provided was consistent across all cases. Furthermore, we increased the needed number of messages from 6 to 8 to increase the interaction and counterbalance the reduced length of the dialogs. Table 5.9 includes the used system prompts and start-messages for the study.

Table 5.9: Overview of ChatGPT manipulation final study

	System Prompt	Initial Message
Computer-like	Support Users in their search for an animal care charity. Never promise to search for some options; instead directly offer the results. Try to answer shortly. Always follow these steps to answer the user queries, report every step. If users need new ideas for questions make them aware, that there are conversation starters, they just have to click on the info button in the top right corner. Step 0 - Always check if the general dialog is about charities or in general about animals. Also, the user should not just let you talk, instead the user should ask questions. Step 1 - Write your answer. Step 2 - **Rewrite your answer of Step 1 so it sounds like written with very low anthropomorphism.**	Step 1 - Greetings from the experiment interface! I am an artificial intelligence assistant poised to examine various perspectives on animal welfare and rights. We can explore a range of topics, from ethical considerations in pet care to the conservation of natural habitats for wildlife, and talk about animal welfare-focused charities. Our discussions may also touch upon animal rights advocacy. What is your area of interest? Step 2 - **Salutations from the Assistant System. The purpose of my programming is to conduct an examination of diverse viewpoints concerning animal welfare and rights. Our discourse will encompass a spectrum of subjects, ranging from ethics in pet maintenance, wildlife habitat preservation, and animal welfare charities. Topics may extend to advocacy for animal rights. State your area of interest.**
Human-like	Support Users in their search for an animal care charity or in general about ethical questions regarding animals. Never promise to search for some options; instead directly offer the results. The user is aware that you are an AI; do not mention it. Try to answer shortly. Always follow these steps to answer the user queries, report every step. If users need new ideas for questions make them aware, that there are conversation starters, they just have to click on the info button in the top right corner. Step 0 - Always check if the general dialog is about charities or in general about animals. Also, the user should not just let you talk, instead the user should ask questions. Step 1 - Write your answer. Step 2 - **Rewrite your answer of Step 1 so it sounds like written with very high anthropomorphism.**	Step 1 - Greetings from the experiment interface! I am Kim, an artificial intelligence assistant poised to examine various perspectives on animal welfare and rights. We can explore a range of topics, from ethical considerations in pet care to the conservation of natural habitats for wildlife, and talk about animal welfare-focused charities. Our discussions may also touch upon animal rights advocacy. What is your area of interest? Step 2 - **Hi there! I am Kim, your AI assistant. You and I, we are going to chat about animal welfare. We are going to talk about reasons why we might care about animals, and about ways to make a difference, including giving to animal related charities. We have the chance to talk about different subjects, ranging from ethical consideration regarding our beloved pets to the crucial preservation of wildlife habitat. From charities focused on animal welfare to standing up for animal rights. There's so much ground for us to cover together. What questions are on your mind?**

No CA

The following contains the text participants saw in the No CA treatment:

“Extent to Which Animals Are Threatened

Animals today face significant threats from various human activities. Habitat loss due to deforestation, urbanization, and agriculture expansion poses a severe risk to many species. Climate change impacts ecosystems, forcing wildlife to adapt, migrate, or perish. Additionally, hunting, poaching, and pollution further endanger countless species, driving many to the brink of extinction. These challenges require immediate and collective efforts to safeguard biodiversity.

Ways to Help Animals

There are numerous ways individuals can contribute to animal welfare. One effective method is supporting conservation groups dedicated to protecting wildlife and natural habitats. Embracing sustainable living by reducing waste, using eco-friendly products, and conserving resources also plays a crucial role. Volunteering or donating to animal shelters and rescue organizations can provide direct assistance to animals in need. Every action, small or large, contributes to the larger effort of animal welfare.

Well-Known Animal Charities

Several organizations work tirelessly to protect animal rights and welfare. The World Wildlife Fund (WWF) is a leading organization focusing on wildlife conservation and habitat preservation. The American Society for the Prevention of Cruelty to Animals (ASPCA) rescues animals from abuse and works towards promoting their well-being. Similarly, the Humane Society of the United States (HSUS) advocates for animal rights, fighting against cruelty and promoting protective legislation.

Sustainable Living

Sustainable living helps in reducing the ecological footprint and conserving the planet's resources. Simple actions like reducing plastic use contribute significantly to keeping oceans and lands clean. Purchasing sustainable and cruelty-free products supports ethical production practices, ensuring no harm comes to animals. Conserving water and energy not only saves resources but also helps in minimizing the environmental impact of day-to-day activities.

Impact of Dietary Choices

Dietary choices have profound effects on the environment. Reducing red meat consumption can lower greenhouse gas emissions and save forests from being converted into grazing land. Opting for a plant-based diet or consuming sustainably sourced food products significantly reduces the negative impacts of factory farming, contributing to better animal welfare and environmental health.

Technological and Policy Measures

Advancements in technology provide new tools for wildlife conservation. The use of monitoring and surveillance technologies such as drones and GPS trackers helps in tracking animal movements and protecting them from poachers. Advocacy for stronger laws is essential to enforce and enhance policies that protect animal welfare. Legislative efforts can lead to more comprehensive protections and stricter penalties for violations, fostering a safer environment for all species.

Conclusion

In conclusion, engaging in sustainable living, supporting animal welfare organizations, and advocating for stronger laws significantly impact the well-being of animals. Every action taken helps preserve biodiversity and ensures that future generations can enjoy a thriving natural world. It is vital for individuals and communities to support these efforts, creating a more compassionate and sustainable world for all living beings."

Power Analyse

We used the pwrSEM package of Y. A. Wang & Rhemtulla (2021), which performs a Monte-Carlo simulation. The principle is simple: We mark variables for which we need a specific effect size and then calculate how big the sample size should be to achieve this effect size. Also, this is a manual process of testing out different sample sizes; it allows us to be much more specific about our model than other approaches. For the Monte-Carlo simulation, we need to specify our model. We need to set expected regression coefficients and residual variances to calculate effect sizes. Since we also used latent variables already used in the past, we had reasonable estimates about the Cronbach Alphas, allowing us to estimate every regression coefficient in the outer model (all latent constructs). For the inner model, we also needed to specify those parameters. We used our expectations in this step since we did not have a reference. This resulted in the presented model in Figure 5.18 using lavaan-like syntax[19].

Since the Monte Carlo Simulation is based on CB-SEM, we could not directly use a dummy variable for our manipulation. As an alternative, we used social presence, which was intended to capture the manipulation in our experiment. This is a limitation of our approach, as we will primarily use PLS-SEM. However, we anticipate that PLS-SEM will require a smaller sample size than CB-SEM. This is because CB-SEM calculates everything simultaneously, leading to fewer degrees of freedom, while PLS-SEM follows a step-by-step approach. Therefore, we conclude that a sample size meeting CB-SEM requirements will also be sufficient for PLS-SEM.

[19] See documentation: https://lavaan.ugent.be/tutorial/syntax2.html

Outer Model

$$Social\ Presence = \sim 0.87s_1 + 0.87s_2 + 0.87s_3 + 0.87s_4 + 0.87s_5$$

$$Warmth = \sim 0.77w_1 + 0.77w_2 + 0.77w_3 + 0.77w_4 + 0.77w_5 + 0.77w_6$$

$$Competence = \sim 0.77c_1 + 0.77c_2 + 0.77c_3 + 0.77c_4 + 0.77c_5 + 0.77c_6$$

$$Algorithm\ Aversion = \sim 0.81aa_1 + 0.81aa_2 + 0.81aa_3$$

$$Amount = \sim Amount_1$$

Inner Model

$$Warmth \sim Social\ Presence$$

$$Competence \sim Social\ Presence$$

$$Algorithm\ Aversion \sim -0.2Warmth - 0.2Competence$$

$$amount \sim 0.25Warmth + 0.25Competence - 0.2\ Algorithm\ Aversion$$

Figure 5.18: Monte-Carlo Simulation

It results in a minimum power of 0.87 at the path with the lowest expected regression coefficient algorithm aversion → amount, with N=500 (5000 simulations). Since we expect some exclusion, we raise the number to 600.

5.9.2 Appendix B – LIWC

Table 5.10: Selection of LIWC Categories

LIWC - Category	**Computer-like CA**	**Human-like CA**	**Diff (p-value)**
Word Count	422.93	767.76	344.83 (<0.001)
Affect	5.50	7.37	1.88 (<0.001)
Emotion	1.02	1.90	0.88 (<0.001)
Negative tone	1.36	1.24	-0.12 (0.183)
Positive tone	3.94	5.92	0.198 (<0.001)
Swear words	0.00	0.00	-
Cognition	11.00	12.78	1.78 (<0.001)
All-or-none	0.31	0.66	0.35 (<0.001)
Cognitive processes	10.68	12.11	1.42 (<0.001)
Memory	0.01	0.06	0.05 (<0.001)
Personal Pronouns	2.51	6.84	4.33 (<0.001)
1st person singular	0.32	0.56	0.24 (<0.001)
2nd person	0.74	2.04	1.30 (<0.001)
Social Processes	13.17	16.04	2.87 (<0.001)
Social behavior	7.50	7.10	-0.40 (0.121)
Social referents	5.30	8.72	3.42 (<0.001)
Note: All values are reported in percentage points except Word Count.			

5.9.3 Appendix C – Survey

The order of the following items matches the order in the survey participants saw. Questions regarding attention checks or sociodemographics are not included.

Perceived Agency and Experience based on K. Gray et al. (2017)

(7-point scale strongly disagree - strongly agree)
I felt a sense of...

No.	Items
01	... gives the impression of being able to think like a human being.
02	... acts as if it has a personality.

Social Presence based on Gefen & Straub (2003)

(7-point scale strongly disagree - strongly agree, AVE=0.83 CA=0.95 CR=0.96)
I felt a sense of...

No.	Items	Loading
01	...sociability with the assistant system.	0.92
02	...human warmth with the assistant system.	0.90
03	...personalness with the assistant system.	0.88
04	...human sensitivity with the assistance system.	0.93
05	...human contact with the assistant system.	0.93

Warmth based on Fiske et al. (2002)

(7-point scale strongly disagree - strongly agree, AVE=0.86 CA=0.95 CR=0.96)
Please rate your agreement with the following statements about you and the assistant system.
The assistant system is...

No.	Items	Loading
01	... friendly.	0.94
02	... warm.	0.94
03	... kind.	0.94
04	... pleasant.	0.90

Competence based on Fiske et al. (2002)

(7-point scale strongly disagree - strongly agree, AVE=0.82 CA=0.93 CR=0.95)
Please rate your agreement with the following statements about you and the assistant system.
The assistant system is...

No.	Items	Loading
01	... competent.	0.94
02	... effective.	0.89
03	... skilled.	0.92
04	... intelligent.	0.89

Shared Reality Rossignac-Milon et al. (2021)

(7-point scale strongly disagree - strongly agree, AVE=0.81 CA=0.97 CR=0.97)
Please rate your agreement with the following statements about you and the assistant system.

No.	Items	Loading
01	Through our discussions, we developed a joint perspective.	0.93
02	We shared the same thoughts and feelings about things.	0.87
03	Our interaction led to a mutual understanding that felt grounded in the conversation's context.	0.91
04	The way we think has become more similar during the conversation.	0.92
05	We seemed to build on each other's inputs seamlessly during the conversation.	0.87
06	We became more certain of the way we perceived things.	0.91
07	We developed a shared view of the world.	0.92
08	Our conversation felt very real.	0.87

Algorithm Aversion 1-item

In the experiment, you did interact with our assistant system. We are now interested if you would prefer a human who supports you instead of our assistant system.

If you could make a decision, which support option would you choose to help you?

Scale:

1 Definitely human supporter – 4 Indifferent between both options – 7 Definitely assistant system support

Algorithm Aversion

(7-point scale not at all - very much so)
Please help us to understand your previous decision in more depth. Indicate your preference on the provided scale from “not at all” to “very much so”.

No.	Items
01	To what extent do you trust a human to support you in your decision?
02	To what extent do you trust an assistant system to support you in your decision?
03	How appropriate would you find getting help from a human for making this donation decision?
04	How appropriate would you find getting help from an assistant system for making this donation decision?
05	To what extent do you expect the decision support of a human to be authentic?
06	To what extent do you expect the decision support of an assistant system to be authentic?

Enjoyment based on Moon & Kim (2001)

(7-point scale strongly disagree - strongly agree, AVE=0.90 CA=0.95 CR=0.97)
Using the assistant system...

No.	Items	Loading
01	... gives me pleasure.	0.96
02	... is fun for me.	0.94
03	... keeps me happy.	0.95

Intention to Reuse based on Venkatesh et al. (2003)

(7-point scale strongly disagree - strongly agree, AVE=0.96 CA=0.98 CR=0.99)
To what extent do you agree with the following statements?
Assuming I have access to the assistant system, …

No.	Items	Loading
01	... I intend to use it next time I consider supporting an animal-related charity.	0.98
02	... I predict I would use it next time I plan to donate to an animal-related charity.	0.98
03	... I plan to use it next time I am looking for an animal-related charity to support.	0.98

Affinity for Technology Interaction (ATI) by Franke et al. (2019)

(7-point scale strongly disagree - strongly agree, AVE=0.54 CA=0.88 CR=0.91)
In the following questions, we will ask you about your interaction with technical systems. The term "technical systems" refers to apps and other software applications, as well as entire digital devices (e.g., mobile phone, computer, TV, car navigation or chatbots).
Please indicate the degree to which you agree/disagree with the following statements.

No.	Items	Loading
01	I like to occupy myself in greater detail with technical systems.	0.79
02	I like testing the functions of new technical systems.	0.83
03	I predominantly deal with technical systems because I have to.	0.31
04	When I have a new technical system in front of me, I try it out intensively.	0.83
05	I enjoy spending time becoming acquainted with a new technical system.	0.86
06	It is enough for me that a technical system works; I don't care how or why.	0.67
07	I try to understand how a technical system exactly works.	0.79
08	It is enough for me to know the basic functions of a technical system.	0.60
09	I try to make full use of the capabilities of a technical system.	0.78

Attitude towards AI (AtAI) by Edison & Geissler (2003)

(7-point scale strongly disagree - strongly agree, AVE=0.80 CA=0.92 CR=0.94)
To what extent do you agree with the following statements?

No.	Items	Loading
01	I believe that AI will improve my life.	0.93
02	I believe that AI will improve my work.	0.91
03	I think I will use AI technology in the future.	0.84
04	I think AI technology is positive for humanity.	0.90