

Technical Autonomy in Human-Machine Collaboration:

An Exploration of Human Responses to Advanced Technology

DISSERTATION
of the University of St.Gallen,
School of Management,
Economics, Law, Social Sciences
and International Affairs
to obtain the title of
Doctor of Philosophy in Management

submitted by

Ruth Frischknecht

from

Schwellbrunn (Appenzell Ausserrhoden)

Approved on the application of

Prof. Dr. Kuno Schedler

and

Prof. Dr. Tina Kiefer

Dissertation Nr. 5218

Difo-Druck GmbH, Untersiemau / 2022

The University of St.Gallen, School of Management, Economics, Law, Social Sciences, International Affairs and Computer Science, hereby consents to the printing of the present dissertation, without hereby expressing any opinion on the views herein expressed.

St.Gallen, May 24, 2022

The President:

Prof. Dr. Bernhard Ehrenzeller

Preface

Throughout the duration of this project, I have been fortunate to meet people who have wished me well and supported, challenged, and encouraged me. Special thanks goes to my supervisor, Prof. Dr. Kuno Schedler. You made this research possible and gave me the freedom I needed, from conception to implementation. Thank you for creating an atmosphere in which I could develop my ideas to their fullest extent, and I am grateful for your advice. I have learned a lot from you.

I would also like to thank my second supervisor, Prof. Dr. Tina Kiefer, for always taking time to work with me, and for her profound expertise, which helped me with every challenge a researcher might face. Your unique ability to always ask exactly the right question has improved this project significantly.

At the beginning of this project, I was lucky enough to meet someone who shared my interests and academic passion. Therefore, a very special thanks goes to Prof. Dr. Monika Simmler. You not only dared to write a paper with a complete beginner (me), you supported and encouraged me throughout this endeavor. It was the many discussions with you that put this piece of work on solid ground. You made all the difference!

I also want to thank my two office mates and friends, Drs. Tobias Tomczak and Aline Tannò. You have both been such an asset to me. My friends from outside university have also played a pivotal role. Thank you for the good company, for every single light-hearted hour with you, and for keeping me grounded.

My heartfelt thanks go to my family. To my parents, Elsbeth and Stefan, and my brother Urs, for showing me the value of common sense, dedication, and compassion. And to Marco, your serenity and wit have been a source of strength for me. To be able to count on so much support and love is a rare and precious gift. Thank you for always believing in me.

St. Gallen, October 2021

Ruth Frischknecht

Summary

Based on the assumptions that technology is becoming increasingly autonomous and new human-machine constellations are emerging, this thesis explores the question of how humans respond to such increasingly autonomous technical systems. Although many tasks are accomplished in close collaboration between humans and machines, existing research insufficiently incorporates the user's (or collaborator's respectively) perspective.

Three articles address the overarching research question. The first article introduces the concept of technical autonomy, derives five levels of this property, and distinguishes it from the notion of automation. For exploring human responses, two online experiments are conducted. The first experiment tests the effect of technical autonomy on general social perceptions ($N = 288$), and the second experiment analyzes the effect of technical autonomy in organizational decision-making scenarios ($N = 341$).

The thesis presents evidence that the level of autonomy affects how people respond to technical systems. It suggests that people perceive technical systems that are high in autonomy as having more agency and competence, and are more willing to rely on them than systems operating on lower levels of autonomy. Third, this effect is particularly evident when technical systems are described as adaptive or both, adaptive and open.

The results of this research will inform both research and practice. For the former, the results suggest that technical autonomy is a relevant variable in human-machine collaboration and this technical property should be further incorporated into research in general and theory-building. For the latter, the findings suggest that autonomous systems should be implemented with caution.

Zusammenfassung

Ausgehend von den Annahmen, dass Technik immer autonomer wird und gleichzeitig neue Mensch-Maschine-Konstellationen entstehen, geht die vorliegende Thesis der Frage nach, wie Menschen auf zunehmend autonome technische Systeme reagieren. Obwohl viele Aufgaben in enger Zusammenarbeit zwischen Menschen und Maschinen erledigt werden, bezieht die bestehende Forschung die Perspektive der Nutzerin bzw. des Nutzers nur unzureichend ein.

Insgesamt drei Artikel befassen sich mit dieser Forschungsfrage. Der erste Artikel führt das Konzept der technischen Autonomie ein, leitet fünf Stufen dieser Eigenschaft ab und grenzt sie vom Begriff der Automatisierung ab. Um die menschliche Sicht abzubilden, wurden danach zwei Online-Experimente durchgeführt. Das erste Experiment untersucht die Auswirkungen technischer Autonomie auf die allgemeine soziale Wahrnehmung ($N = 288$), und das zweite Experiment analysiert die Auswirkungen technischer Autonomie in Entscheidungssituationen ($N = 341$).

Die Ergebnisse deuten darauf hin, dass sich sowohl die soziale Wahrnehmung als auf die Bereitschaft auf ein technisches System zurückzugreifen in Abhängigkeit des Autonomiegrades verändern. Menschen scheinen technische Systeme mit einem hohen Autonomiegrad als handlungsfähiger, kompetenter und akzeptabler einzuschätzen als Systeme mit einem niedrigeren Autonomiegrad. Dieser Effekt zeigt sich, wenn Systeme als lernfähig beziehungsweise als lernfähig und offen beschreiben werden.

Für die Wissenschaft bedeuten die Ergebnisse, dass technische Autonomie eine relevante Variable in der Mensch-Maschine-Kollaboration ist und diese technische Eigenschaft weiter in die Forschung im Allgemeinen und die Theoriebildung im Besonderen einbezogen werden sollte. Für die Praxis bedeuten die Ergebnisse, dass die Implementierung autonomer Systeme umsichtig und sorgfältig erfolgen soll.

Content

Preface.....	iii
Summary	iv
Zusammenfassung.....	v
Content	vi
List of Figures	ix
List of Tables.....	x
Abbreviations	xi
1 Introduction	1
1.1 Autonomous Technologies.....	1
1.2 Human-Machine Collaboration.....	2
1.3 Research Gap and Research Question	4
2 Structure of the Thesis.....	6
2.1 Article I: A taxonomy of human-machine collaboration: Capturing automation and technical autonomy	6
2.2 Article II: A social cognition perspective on technical autonomy	7
2.3 Article III: Technical autonomy in human-machine collaboration.....	8
3 Research Approach.....	10
4 References	12
Article I - A taxonomy of human-machine collaboration: capturing automation and technical autonomy.....	18
1 Introduction	19
2 Methodical framework.....	21
3 Theoretical Background	22
3.1 Conceptualizing socio-technical phenomena	22
3.2 Synthesis.....	25
4 Taxonomy	26
4.1 The axes: automation and autonomy.....	26
4.2 Levels of automation	29
4.2.1 Level 1: Offers decision.....	30

4.2.2	Level 2: Executes with human approval	30
4.2.3	Level 3: Executes if no human vetoes	31
4.2.4	Level 4: Executes and then informs	31
4.2.5	Level 5: Executes fully automated.....	31
4.3	Levels of autonomy	32
4.3.1	Level 1: Deterministic system	35
4.3.2	Level 2: Non-transparent system	35
4.3.3	Level 3: Indetermined system.....	36
4.3.4	Level 4: Adaptable system	36
4.3.5	Level 5: Open system	37
5	Discussion	38
5.1	Limitations	41
5.2	Avenues for further research	41
6	Conclusion	42
7	References	44
	Article II – A social cognition perspective on technical autonomy	50
1	Introduction	51
2	The social perception of technology	53
2.1	The present study.....	56
3	Method	59
3.1	Participants	59
3.2	Materials.....	59
3.2.1.	Manipulation.	59
3.2.2.	Dependent variables.	61
3.2.3.	Control variables.	61
3.2.4.	Manipulation and attention checks.	62
3.3	Procedure	62
4	Results.....	63
4.1	Preliminary Analyses	63
4.2	Effects of technical autonomy	64
4.3	Contrasting technical systems with humans	67
5	Discussion	68
6	Conclusion	73

7	References	74
	Article III – Technical autonomy in human-machine collaboration: Effects on the acceptance of advice and delegation of decisions.....	79
1	Introduction	80
2	Theoretical Background	82
2.1	Algorithm aversion and algorithm appreciation	82
2.2	Autonomous technologies	84
2.3	Hypotheses Development	85
3	Method	87
3.1	Design and independent variables.....	87
3.2	Dependent variables.....	90
3.3	Covariates and additional variables.....	90
3.4	Procedure	91
3.5	Sample	92
4	Results.....	92
4.1	Preliminary analyses	92
4.2	Hypotheses testing.....	95
4.3	Additional analyses	97
5	Discussion	98
6	Conclusion	102
	Appendix A.....	103
7	References	104
1	General Discussion.....	109
1.1	Conceptual Contribution	109
1.2	Empirical Contributions.....	110
1.3	Overall Contribution	112
2	Practical Implications	112
3	Limitations	114
4	Areas for Further Research.....	116
5	References	118
	Appendix B: A note about the linearity of the levels of technical autonomy .	121

List of Figures

Figure 1: Taxonomy of automation and technical autonomy	38
Figure 2: Mean ascriptions of social cognition and mind perception.....	66
Figure 3: Mean ascriptions of social cognition and mind perception including the human control condition.....	67
Figure 4: Sample vignette showing the scenario.....	89
Figure 5: Mean willingness to accept advice	96
Figure 6: Mean willingness to delegate a decision	97

List of Tables

Table 1: Information about and publication status of article I.	7
Table 2: Information about and publication status of article II.	8
Table 3: Information about and publication status of article III.....	9
Table 4: Levels of automation with their main features.	30
Table 5: Dimensions of technical autonomy with assessment questions.	34
Table 6: Levels of autonomy with an explanation of the main features.	35
Table 7: Five levels of technical autonomy and main characteristics.....	58
Table 8: Descriptive statistics of the dependent variables.	64
Table 9: Correlations among dependent variables.....	65
Table 10: Descriptive statistics of the dependent variables.	94
Table 11: Technical system descriptions provided in the vignettes, ...	103

Abbreviations

ANOVA	Analysis of Variance
CI	Confidence Interval
F	F-value
M	Mean
MANOVA	Multivariate Analysis of Variance
M-Turk	Amazon's Mechanical Turk
NHTSA	National Highway Traffic Safety Administration
p	p-value
SD	Standard Deviation
SE	Standard Error

1 Introduction

In the last several decades, rapid and fundamental technical development has instigated the significant spread of (smart) technical systems capable of assuming an increasing number of (previously human) tasks (Brynjolfsson & McAfee, 2014; Wilson & Daugherty, 2018). There are numerous examples of this. In health care, an app for detecting skin cancer can now achieve the same level of accuracy as a dermatologist (Haenssle et al., 2018). In law enforcement, complex models are used to determine the likelihood of offender recidivism (Dressel & Farid, 2018). Advanced chatbots and voice assistants can not only communicate with users but are now also able to recognize human emotions based on voice and tone (Wilson & Daugherty, 2018). This list is, of course, not exhaustive, and likely to extended almost infinitely. What these examples show is that our society is entrusting technical systems with a steadily growing number of tasks that would have been considered impossible to automate just a few years ago (Brynjolfsson & McAfee, 2014; Duan, Edwards, & Dwivedi, 2019). This thesis scientifically captures the implications of these newly evolved and rapidly advancing technical systems on human-machine interaction, establishing a fundamental understanding of how increasingly intelligent machines affect human responses. The following sections outline the basic premises upon which this thesis is based. Then, the research question is stated, project structure explained, and key methodological considerations addressed.

1.1 Autonomous Technologies

At the heart of the advances described above lie new capabilities and advanced technical capacity that enable technical systems to complete highly complex tasks in the first place. Often, the so-called intelligence of these systems is based on their ability to access, process, and analyze large amounts of data (Duan et al., 2019; Kaplan, 2015; Russell & Norvig, 2016) and engage in machine learning (Jordan & Mitchell, 2015; Russell & Norvig, 2016). Sometimes they are capable of interacting with other technical systems (Floridi & Sanders, 2004; Sartor &

Omicini, 2016), thereby creating multi-agent systems that exceed the capacity of a single agent (Thurmel, 2015; Weyer & Reineke, 2012). Because of these capabilities, advanced technologies appear to be intelligent and thus are commonly referred to as artificial intelligence (Russell & Norvig, 2016). As they can learn and adjust their behavior, these machines (or programs) are referred to as (technical) *agents* (Russell & Norvig, 2016; Seeber, Waizenegger, et al., 2020), underscoring that advanced technologies exert agency (Rammert, 2008). They are able to respond to situations without external intervention and thus act independently, at least to a certain extent (Floridi & Sanders, 2004; Russell & Norvig, 2016). The ability to act independently means that these technical agents are *autonomous* (Castelfranchi & Falcone, 2004; de Visser, Pak, & Shaw, 2018; Seeber, Bittner, et al., 2020). Thus, a recognition of *technical autonomy is essential in the digital age*.

Agency in technical systems is conceived of as graduated and consisting of different levels of activity: passive, semi-active, reactive, proactive, and cooperative (Rammert, 2008, 2011). Because technical autonomy is closely interwoven with agency, it is useful to understand not only agency but also autonomy as graduated and comprised of different levels (a more detailed derivation is provided in the first article.) Thus, this thesis conceptualizes *technical autonomy as gradual and consisting of different levels*.

1.2 Human-Machine Collaboration

As mentioned above, autonomous technical systems are increasingly capable of performing complex tasks; many potential combinations of humans and machines have emerged for such purposes. In some cases, autonomous technical systems assume complete control, without the need for human intervention (Hoddinghaus, Sondern, & Hertel, 2021; Lee, 2018). However, in most cases, machines cannot suddenly replace people. Rather, it is likely that more and more tasks will be performed in close cooperation between humans and (autonomous) technical systems (Duan et al., 2019). For example,

autonomous technical systems are currently being used to inform human decision-makers, suggest alternatives, and even recommend optimized choices (Jarrahi, 2018; Makarius, Mukherjee, Fox, & Fox, 2020). Thus, humans and machines can establish symbiosis (Jarrahi, 2018), form relationships (de Visser et al., 2018) and be described as teammates (McNeese, Demir, Cooke, & Myers, 2018; Seeber, Bittner, et al., 2020). Consequently, with the spread of autonomous technical systems, *new human-machine constellations have emerged*.

Especially the close collaboration between humans and machines promises to augment organizational decision-making (Raisch & Krakowski, 2021; Wilson & Daugherty, 2018), foster innovation (Benbya, Pachidi, & Jarvenpaa, 2021; Brynjolfsson & McAfee, 2012; Raisch & Krakowski, 2021), enhance financial performance (Davenport & Ronanki, 2018), create strategic advantages (Wilson & Daugherty, 2018), and improve service quality (Davenport & Ronanki, 2018; Wilson & Daugherty, 2018). However, introducing autonomous technical systems also entails risks. There are concerns that the spread of artificial intelligence will eliminate jobs (Brynjolfsson & McAfee, 2012; Russell & Norvig, 2016), lead to a loss of human control and oversight (Brynjolfsson & McAfee, 2014; Diakopoulos, 2016; Russell & Norvig, 2016; Zarsky, 2016), increase discrimination (O'Neil, 2016), and enhance a general aversion to or mistrust of such applications (Burton, Stein, & Jensen, 2020; Dietvorst, Simmons, & Massey, 2015). From this, it follows that in organizations, the success or failure of increasingly autonomous technical systems will largely depend on how people (i.e., their users) respond to them. After all, in most cases, it will be people who decide whether to follow the advice of a machine or rely on a technical system. Thus, for organizations to realize their potential while still avoiding the undesirable consequences of the presence of autonomous technical systems, the impact such machines have on people must be thoroughly investigated. *Understanding the effect of increasing technical autonomy on human response is therefore crucial in the digital age.*

1.3 Research Gap and Research Question

Despite the spread of advanced technology and its relevance to organizations, research in the field of human-machine collaboration has not systematically accounted for the effects of technical autonomy. Rather, technical functioning is currently regarded as a constant (Sundar, 2020), meaning that autonomous technology is said to elicit more or less the same responses as simpler communications technology (Nass & Moon, 2000; Reeves & Nass, 1996). However, with the emergence of advanced machines capable of acting autonomously, the notion that technical functioning does not matter for social responses is no longer likely to be true (see Schuetz & Venkatesh, 2020; Sundar, 2020). Traditionally, technical systems have been understood as purely instrumental tools that function passively, in contrast with humans who are able to act (actively) (Rammert, 2011; Rammert & Schulz-Schaeffer, 2002). However, advanced autonomous technologies blur the boundaries between human action and technical functioning (Rammert & Schulz-Schaeffer, 2002). Consequently, long-held and broadly established assumptions about human-technology dualism are in need of refinement (de Visser et al., 2018; Schuetz & Venkatesh, 2020; Sundar, 2020). Moreover, a strict division between technological functioning and human action prevents the emergence of a finer-grained understanding of advanced technology (Rammert, 2011) that adequately reflects current technical developments.

The majority of research on human responses to and perceptions of technology have tended to compare humans and machines (see for example Cadario, Longoni, & Morewedge, 2021; Gray, Gray, & Wegner, 2007; Logg, Minson, & Moore, 2019) and have not systematically accounted for the level of autonomy on which the particular technical system operates. This is true despite technical capabilities often being discussed as a significant driver of the human response to such machines (Glikson & Woolley, 2020). Moreover, although it is important to deconstruct human ascriptions, most research on technical autonomy has been approached from a technical side, conducted by programmers and

engineers and usually with the goal of improving or optimizing performance (Glikson & Woolley, 2020; Rahwan et al., 2019; Raisch & Krakowski, 2021).

From this, it follows that existing research on human-machine collaboration has not sufficiently captured recent technical developments increasing technical autonomy, and the effect of such increases on the human response is not well-understood. Thus, this work endeavored to answer the following research question.

What is the effect of increasing autonomy on human responses to technical systems?

Answering this question will contribute to a deeper understanding of technical autonomy per se and help to clarify the impact of increasing levels of technical autonomy in the digital age. It is important to note that it is as yet unclear if the human response to increasingly autonomous technical systems changes at all. It may be that the amount of technical autonomy plays no role and thus does not warrant further theorizing. From a practical point of view, the results of this work will help to inform organizations regarding how humans collaborate with increasingly autonomous technical systems and how those systems and their outputs are perceived. Based on these insights, organizations will be better positioned to develop educational programs and adjust guidelines related to system use.

2 Structure of the Thesis

In this research, a three-part approach was applied to explore the human response to increasingly autonomous technology. The first article described and deconstructed technical autonomy and established the foundation for its empirical investigation. The subsequent articles analyzed various human responses to autonomous technologies. As “response” is a very broad term, it was also further described and analyzed. The second article investigated the impact of (increasingly) autonomous technical systems on social perception, a broad and fundamental set of social responses to different sorts of actors (i.e., other groups of humans) (Fiske, Cuddy, & Glick, 2007), organizations (Aaker, Vohs, & Mogilner, 2010), and robots (Carpinella, Wyman, Perez, & Stroessner, 2017). More narrowly, as analyzed in the third article, response was understood as a willingness to accept advice from or delegate decisions to recommender systems operating on various levels of autonomy, as described in two scenarios representing organizational decision-making. In this sense, response meant the level of reliance on more- or less-autonomous technical systems.

2.1 Article I: A taxonomy of human–machine collaboration: Capturing automation and technical autonomy

The first part of this project was conceptual in nature. In the literature review, the concept of technical autonomy was clarified and distinguished from automation. In the existing body of research, automation and (technical) autonomy have often been used interchangeably (Vagia, Transeth, & Fjerdingen, 2016). Autonomy has even been regarded as the highest level of automation (see e.g. Sheridan & Verplank, 1978; Vagia et al., 2016). In the article, it is argued that technical autonomy consists of different levels and thus is conceptualized as a graduated property. The derivations and descriptions of the various stages are well-established in the field of automation (see e.g. Parasuraman, Sheridan, & Wickens, 2000; Sheridan & Verplank, 1978). For technical autonomy, no similar approach has thus far been developed.

However, there have been attempts to determine the dimensions (Onnasch, Maier, & Jürgensohn, 2016) and capabilities (Beer, Fisk, & Rogers, 2014) that constitute technical autonomy. The taxonomy includes these ideas and considerations regarding the different levels of autonomy.

As such, the first article served as a foundation for the subsequent empirical articles, which further investigated the consequences of technologies operating at increasing levels of technical autonomy. In sum, the independent variable and its levels were defined, explained, and related to other relevant concepts, allowing a more differentiated understanding of advanced technologies to emerge.

Table 1: Information about and publication status of article I.

A taxonomy of human–machine collaboration: Capturing automation and technical autonomy	
Authors	Monika Simmler & Ruth Frischknecht
Journal	AI & Society
Impact Factor	1.61
Status	Published

2.2 Article II: A social cognition perspective on technical autonomy

The second article was the first to test human responses to the five levels of autonomy described in the taxonomy. It employed a social cognition perspective, meaning it investigated the impacts of increasing levels of technical autonomy on ascription of the social cognition variables of warmth and competence (Fiske et al., 2007) and mind perception variables of experience and agency (Gray et al., 2007).

Both social cognition and mind perception have been found to be determinants of human behavior and are regarded as foundational for a multitude of social responses such as trust (Hu, Lu, Pan, Gong, & Yang, 2021; Kulms & Kopp, 2018) and the ascription of responsibility (Gray et al., 2007; Waytz, Gray, Epley, &

Wegner, 2010). Both trust and responsibility are especially important in organizational settings.

Studies investigating how technologies are socially perceived have usually compared technical systems or robots to humans (e. g. Gray et al., 2007). It has been shown repeatedly that technical systems are ascribed a moderate amount of competence and agency, but almost no experience or warmth (Pan, Croft, & Niemeyer, 2018; Stafford, MacDonald, Jayawardena, Wegner, & Broadbent, 2014). Yet, whether and how increasing levels of autonomy alter such social perceptions has yet to be thoroughly investigated. Additionally, it is not entirely clear if (increasing) technical autonomy has an effect on such ascriptions at all. Thus, this article addressed a basic question and intentionally refrained from measuring more specific responses or considering the context in which the technical system is deployed.

Table 2: Information about and publication status of article II.

A social cognition perspective on technical autonomy	
Author	Ruth Frischknecht
Journal	Computers in Human Behavior
Impact Factor	6.829
Status	Published

2.3 Article III: Technical autonomy in human-machine collaboration

Because the second article investigated general social perceptions, the third focused on more specific responses to technical systems and added context by introducing several decision-making scenarios. Different human-machine constellations have emerged in a variety of organizations, some characterized through joint decisions made by human and machine, such as when a recommender system suggests options and the human decides (Jarrahi, 2018; Raisch & Krakowski, 2021), and others involving technical systems empowered to make choices on their own (Hoddinghaus et al., 2021; Lee, 2018). The third article examined and analyzed such constellations, investigating whether the

level of technical autonomy had an effect on humans' willingness to accept advice from and delegate decisions to such devices, using as examples one investment and one personnel decision.

This is relevant because technical autonomy reduces the transparency, comprehensibility, and traceability of the particular system (Rahwan et al., 2019; Shin & Park, 2019). As people tend to be reluctant to use technical tools to inform their decisions (Dietvorst et al., 2015) and prefer advice from algorithms they consider comprehensible (Cadario et al., 2021), further understanding of the role of technical autonomy is crucial. Ignoring the advice of machines means that their potential to positively impact human decisions (Wilson & Daugherty, 2018) is not being adequately harnessed.

Table 3: Information about and publication status of article III.

Technical autonomy in human-machine collaboration: Effects on the acceptance of advice and delegation of decisions	
Author	Ruth Frischknecht
Journal	International Journal of Human-Computer Interaction
Impact Factor	3.35
Status	Submitted, 30.09.2021

3 Research Approach

Given the newness of the phenomenon under investigation, the work presented herein should be understood as basic research. Basic research is concerned with elementary premises and assumptions and interested in general insights (Bortz & Döring, 2006). Thus, this thesis did not explore specific applications of autonomous technologies, nor did it investigate organizations currently deploying systems with high or low levels of technical autonomy. Rather, it evaluated the fundamental effects of technical autonomy on the human response. Consequently, it raised basic questions such as whether and how technical autonomy might be relevant in human-machine collaboration.

Because the scope of this project was rooted in basic research, the data were collected from experiments. Experiments allow researchers to isolate the effects of one or more independent variables (here, technical autonomy) on one or more dependent variables (here, the willingness to accept advice) (Shadish, Campbell, Campbell, & Cook, 2002; Webster & Sell, 2014). By randomly assigning subjects to different experimental conditions, systematic biases in the sample can be controlled (Alferes, 2012) because it is assumed that individual characteristics (or other external influences) are equally (approximately) distributed across the different experimental conditions (Alferes, 2012; Webster & Sell, 2014). Along with the control of confounding variables, experiments in principle allow for (careful) causal inferences (Shadish et al., 2002; Webster & Sell, 2014). In the present research, this meant that the influence of different levels of autonomy on ascriptions and intentions could be isolated and separately interpreted. Put differently, because the experimental conditions only differed by the level of autonomy, differences in attribution or intention could be connected to the autonomy level (i.e., it was assumed that the autonomy level caused the differences in responses to the system).

Vignettes were used to describe the levels of autonomy. According to Atzmüller and Steiner (2010), a vignette is “a short, carefully constructed description of a person, object or situation, representing a systematic combination of

characteristics” (p. 128). Vignettes are useful for systematically measuring attitudes and intentions (Aguinis & Bradley, 2014). The purpose of vignette-based research is to identify and isolate the influence of factors described in the scenarios (Atzmüller & Steiner, 2010). Typically, in vignette experiments, multiple variables are manipulated simultaneously and varied within and between participants (Auspurg & Hinz, 2015). However, the present research only systematically varied one variable (i.e., technical autonomy). Thus, this thesis only drew upon the basic idea of this method and did not fully implement it. Nevertheless, the case vignettes were useful because they allowed for the manipulation of technical autonomy in a controlled and systematic way.

All participants (N = 288 in the second and N = 341 in the third article) were recruited on Amazon’s Mechanical Turk (M-turk). Although standard samples usually consist of college students (Casler, Bickel, & Hackett, 2013), for the following reasons, using M-turk can be considered an appropriate way to address the research question. M-turk samples have been shown to provide information of similar quality to data collected through more conventional methods (Buhrmester, Kwang, & Gosling, 2011; Casler et al., 2013). Additionally, in both empirical studies, strict inclusion criteria such as an approval rate higher than 95% (Peer, Vosgerau, & Acquisti, 2014) and inclusion of attention checks (Abbey & Meloy, 2017) served to ensure data quality.

In sum, this research project clarified a fundamental question: how (if at all) autonomy affects human responses to technical systems. Different levels of autonomy were derived from the literature review. The effects of these different levels were then tested in two online experiments. This method was suitable because the experiments allowed the researcher to systematically test the effects of increasing levels of autonomy while still ensuring tight control of other possibly confounding factors.

4 References

- 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. doi: 10.1086/651566
- Abbey, J. D., & Meloy, M. G. (2017). Attention by design: Using attention checks to detect inattentive respondents and improve data quality. *Journal of Operations Management*, 53-56(1), 63-70. doi: 10.1016/j.jom.2017.06.001
- Aguinis, H., & Bradley, K. J. (2014). Best Practice Recommendations for Designing and Implementing Experimental Vignette Methodology Studies. *Organizational Research Methods*, 17(4), 351-371. doi: 10.1177/1094428114547952
- Alferes, V. R. (2012). *Methods of randomization in experimental design*. Los Angeles: SAGE.
- Atzmüller, C., & Steiner, P. M. (2010). Experimental Vignette Studies in Survey Research. *Methodology*, 6(3), 128-138. doi: 10.1027/1614-2241/a000014
- Auspurg, K., & Hinz, T. (2015). *Factorial survey experiments*. Los Angeles: Sage.
- Beer, J. M., Fisk, A. D., & Rogers, W. A. (2014). Toward a framework for levels of robot autonomy in human-robot interaction. *Journal of Human-Robot Interaction*, 3(2), 18. doi: 10.5898/2FJHRI.3.2.Beer
- Benbya, H., Pachidi, S., & Jarvenpaa, S. L. (2021). Special Issue Editorial: Artificial Intelligence in Organizations: Implications for Information Systems Research. *Journal of the Association for Information Systems*, 22(2), 281-303. doi: 10.17705/1jais.00662
- Bortz, J., & Döring, N. (2006). Besonderheiten der Evaluationsforschung. In J. Bortz & N. Döring (Eds.), *Forschungsmethoden und Evaluation: für Human- und Sozialwissenschaftler* (pp. 95-136). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Brynjolfsson, E., & McAfee, A. (2012). Winning the Race With Ever-Smarter Machines. *Mit Sloan Management Review*, 53(2), 53-60.
- Brynjolfsson, E., & McAfee, A. (2014). *The Second Machine Age : work, progress, and prosperity in a time of brilliant technologies*. New York: Norton.
- Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? *Perspectives on Psychological Science*, 6(1), 3-5. doi: 10.1177/1745691610393980
- Burton, J. W., Stein, M. K., & Jensen, T. B. (2020). A systematic review of algorithm aversion in augmented decision making. *Journal of Behavioral Decision Making*, 33(2), 220-239. doi: 10.1002/bdm.2155
- Cadario, R., Longoni, C., & Morewedge, C. K. (2021). Understanding, explaining, and utilizing medical artificial intelligence. *Nature Human Behaviour*. doi: 10.1038/s41562-021-01146-0

-
- Carpinella, C. M., Wyman, A. B., Perez, M. A., & Stroessner, S. J. (2017). *The Robotic Social Attributes Scale (RoSAS): Development and Validation*. Paper presented at the Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction, Vienna, Austria.
- Casler, K., Bickel, L., & Hackett, E. (2013). Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing. *Computers in Human Behavior*, 29(6), 2156-2160. doi: 10.1016/j.chb.2013.05.009
- Castelfranchi, C., & Falcone, R. (2004). Founding autonomy: The dialectics between (social) environment and agent's architecture and powers. In M. Nickles, M. Rovatos, & G. Weiss (Eds.), *Agents and computational autonomy: potential, risks, and solutions* (pp. 40-54). Berlin: Springer.
- Davenport, T. H., & Ronanki, R. (2018). Artificial intelligence for the real world: don't start with moon shots. *Harvard Business Review*, 96(1), 108-116.
- de Visser, E. J., Pak, R., & Shaw, T. H. (2018). From 'automation' to 'autonomy': the importance of trust repair in human-machine interaction. *Ergonomics*, 61(10), 1409-1427. doi: 10.1080/00140139.2018.1457725
- Diakopoulos, N. (2016). Accountability in Algorithmic Decision Making. *Communications of the Acm*, 59(2), 56-62. doi: 10.1145/2844110
- 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. doi: 10.1037/xge0000033
- Dressel, J., & Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism. *Science Advances*, 4(1). doi: 10.1126/sciadv.aao5580
- Duan, Y. Q., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of Big Data - evolution, challenges and research agenda. *International Journal of Information Management*, 48, 63-71. doi: 10.1016/j.ijinfomgt.2019.01.021
- 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. doi: 10.1016/j.tics.2006.11.005
- Floridi, L., & Sanders, J. W. (2004). On the morality of artificial agents. *Minds and Machines*, 14(3), 349-379. doi: 10.1023/B:MIND.0000035461.63578.9d
- Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical evidence. *Academy of Management Annals*, 14(2), 627-660. doi: 10.5465/annals.2018.0057
- Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science*, 315(5812), 619-619. doi: 10.1126/science.1134475
- Haenssle, H. A., Fink, C., Schneiderbauer, R., Toberer, F., Buhl, T., Blum, A., . . . Zalaudek, I. (2018). Man against machine: diagnostic performance of a

- deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Ann Oncol*, 29(8), 1836-1842. doi: 10.1093/annonc/mdy166
- Hoddinghaus, M., Sondern, D., & Hertel, G. (2021). The automation of leadership functions: Would people trust decision algorithms? *Computers in Human Behavior*, 116. doi: 10.1016/j.chb.2020.106635
- Hu, Q., Lu, Y. B., Pan, Z., Gong, Y. M., & Yang, Z. L. (2021). Can AI artifacts influence human cognition? The effects of artificial autonomy in intelligent personal assistants. *International Journal of Information Management*, 56, 15. doi: 10.1016/j.ijinfomgt.2020.102250
- Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business Horizons*, 61(4), 577-586. doi: 10.1016/j.bushor.2018.03.007
- Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255-260. doi: 10.1126/science.aaa8415
- Kaplan, J. (2015). *Humans need not apply: a guide to wealth and work in the age of artificial intelligence*. New Haven: Yale University Press.
- Kulms, P., & Kopp, S. (2018). A Social Cognition Perspective on Human-Computer Trust: The Effect of Perceived Warmth and Competence on Trust in Decision-Making With Computers. *Frontiers in Digital Humanities*, 5(14). doi: 10.3389/fdigh.2018.00014
- Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*, 5(1), 2053951718756684. doi: 10.1177/2053951718756684
- 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. doi: 10.1016/j.obhdp.2018.12.005
- Makarius, E. E., Mukherjee, D., Fox, J. D., & Fox, A. K. (2020). Rising with the machines: A sociotechnical framework for bringing artificial intelligence into the organization. *Journal of Business Research*, 120, 262-273. doi: 10.1016/j.jbusres.2020.07.045
- McNeese, N. J., Demir, M., Cooke, N. J., & Myers, C. (2018). Teaming With a Synthetic Teammate: Insights into Human-Autonomy Teaming. *Human Factors*, 60(2), 262-273. doi: 10.1177/0018720817743223
- Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, 56(1), 81-103. doi: 10.1111/0022-4537.00153
- O'Neil, C. (2016). *Weapons of math destruction : how big data increases inequality and threatens democracy*. New York: Crown.

-
- Onnasch, L., Maier, X., & Jürgensohn, T. (2016). *Mensch-Roboter-Interaktion-Eine Taxonomie für alle Anwendungsfälle*: Bundesanstalt für Arbeitsschutz und Arbeitsmedizin Dortmund.
- Pan, M., Croft, E. A., & Niemeyer, G. (2018). Evaluating Social Perception of Human-to-Robot Handovers Using the Robot Social Attributes Scale (RoSAS) *Hri '18: Proceedings of the 2018 Acm/Ieee International Conference on Human-Robot Interaction* (pp. 443-451).
- Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *Ieee Transactions on Systems Man and Cybernetics Part a-Systems and Humans*, 30(3), 286-297. doi: 10.1109/3468.844354
- Peer, E., Vosgerau, J., & Acquisti, A. (2014). Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. *Behavior Research Methods*, 46(4), 1023-1031. doi: 10.3758/s13428-013-0434-y
- Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J.-F., Breazeal, C., . . . Wellman, M. (2019). Machine behaviour. *Nature*, 568(7753), 477-486. doi: 10.1038/s41586-019-1138-y
- Raisch, S., & Krakowski, S. (2021). Artificial Intelligence and Management: The Automation-Augmentation Paradox. *The Academy of Management Review*, 46(1), 192. doi: 10.5465/amr.2018.0072
- Rammert, W. (2008). Where the action is: Distributed agency between humans, machines and programs. Berlin.
- Rammert, W. (2011) Distributed Agency and Advanced Technology Or: How to Analyse Constellations of Collective Inter-Agency. *Technical University Technology Studies Working Papers*. Berlin.
- Rammert, W., & Schulz-Schaeffer, I. (2002) Technik und Handeln - wenn soziales Handeln sich auf menschliches Verhalten und technische Artefakte verteilt. *Technische Universität Berlin Working Papers*. Berlin.
- Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people*: Cambridge university press Cambridge, UK.
- Russell, S. J., & Norvig, P. (2016). *Artificial intelligence : a modern approach*. Boston: Pearson Education.
- Sartor, G., & Omicini, A. (2016). The autonomy of technological systems and responsibilities for their use. In C. Kress, H.-Y. Liu, N. Bhuta, R. Geiss, & S. Beck (Eds.), *Autonomous Weapons Systems: Law, Ethics, Policy* (pp. 39-74). Cambridge: Cambridge University Press.
- Schuetz, S., & Venkatesh, V. (2020). Research Perspectives: The Rise of Human Machines: How Cognitive Computing Systems Challenge Assumptions

- of User-System Interaction. *Journal of the Association for Information Systems*, 21(2), 460-482. doi: 10.17705/1jais.00608
- Seeber, I., Bittner, E., Briggs, R. O., de Vreede, T., de Vreede, G.-J., Elkins, A., . . . Söllner, M. (2020). Machines as teammates: A research agenda on AI in team collaboration. *Information & Management*, 57(2), 103174. doi: <https://doi.org/10.1016/j.im.2019.103174>
- Seeber, I., Waizenegger, L., Seidel, S., Morana, S., Benbasat, I., & Lowry Paul, B. (2020). Collaborating with technology-based autonomous agents: Issues and research opportunities. *Internet Research*, 30(1), 1-18. doi: 10.1108/INTR-12-2019-0503
- Shadish, W. R., Campbell, D. T., Campbell, D. T., & Cook, T. D. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston (Mass.): Houghton Mifflin.
- Sheridan, T. B., & Verplank, W. L. (1978). *Human and computer control in undersea teleoperators*. Massachusetts Institute of Technology, Cambridge: Man-Machine Systems Laboratory Reports.
- Shin, D., & Park, Y. J. (2019). Role of fairness, accountability, and transparency in algorithmic affordance. *Computers in Human Behavior*, 98, 277-284. doi: 10.1016/j.chb.2019.04.019
- Stafford, R. Q., MacDonald, B. A., Jayawardena, C., Wegner, D. M., & Broadbent, E. (2014). Does the Robot Have a Mind? Mind Perception and Attitudes Towards Robots Predict Use of an Eldercare Robot. *International Journal of Social Robotics*, 6(1), 17-32. doi: 10.1007/s12369-013-0186-y
- Sundar, S. S. (2020). Rise of Machine Agency: A Framework for Studying the Psychology of Human-AI Interaction (HAI). *Journal of Computer-Mediated Communication*, 25(1), 74-88. doi: 10.1093/jcmc/zmz026
- Thurmel, S. (2015). The Participatory Turn: A Multidimensional Gradual Agency Concept for Human and Non-human Actors. *Collective Agency and Cooperation in Natural and Artificial Systems: Explanation, Implementation and Simulation*, 122, 45-60. doi: 10.1007/978-3-319-15515-9_3
- Vagia, M., Transeth, A. A., & Fjerdings, S. A. (2016). A literature review on the levels of automation during the years. What are the different taxonomies that have been proposed? *Applied Ergonomics*, 53, 190-202. doi: 10.1016/j.apergo.2015.09.013
- Waytz, A., Gray, K., Epley, N., & Wegner, D. M. (2010). Causes and consequences of mind perception. *Trends in Cognitive Sciences*, 14(8), 383-388. doi: 10.1016/j.tics.2010.05.006
- Webster, M., & Sell, J. (2014). *Laboratory Experiments in the Social Sciences*. San Diego: Elsevier Science & Technology.

- Weyer, J., & Reineke, S. (2012). Creating order in hybrid systems: reflexions on the interaction of man and smart machines. Dortmund.
- Wilson, H. J., & Daugherty, P. R. (2018). Collaborative intelligence: humans and AI are joining forces. *Harvard Business Review*, 96(4), 114.
- Zarsky, T. (2016). The Trouble with Algorithmic Decisions: An Analytic Road Map to Examine Efficiency and Fairness in Automated and Opaque Decision Making. *Science Technology & Human Values*, 41(1), 118-132. doi: 10.1177/0162243915605575

Article I - A taxonomy of human-machine collaboration: capturing automation and technical autonomy

Monika Simmler & Ruth Frischknecht

Abstract

Due to the ongoing advancements in technology, socio-technical collaboration has become increasingly prevalent. This poses challenges in terms of governance and accountability, as well as issues in various other fields. Therefore, it is crucial to familiarize decision-makers and researchers with the core of human-machine collaboration. This study introduces a taxonomy that enables identification of the very nature of human-machine interaction. A literature review has revealed that automation and technical autonomy are main parameters for describing and understanding such interaction. Both aspects must be carefully evaluated, as their increase has potentially far-reaching consequences. Hence, these two concepts comprise the taxonomy's axes. Five levels of automation and five levels of technical autonomy are introduced below, based on the assumption that both automation and autonomy are gradual. The levels of automation were developed from existing approaches; those of autonomy were carefully derived from a review of the literature. The taxonomy's use is also explained, as are its limitations and avenues for further research.

Keywords: *Human-machine collaboration, taxonomy, automation, autonomy*

1 Introduction

In the digital age, human–machine interaction is an essential part of everyday life, such as when pilots fly airplanes, physicians make use of diagnostic programs, workers operate industrial machines, and Uber drivers receive messages about jobs. All of these examples share a commonality: a task is accomplished in collaboration with an advanced machine. In many areas, actions are increasingly the result of such human–machine collaborations (Kirchkamp & Strobel, 2019). Machines are taking over more and more tasks that previously were performed by humans (Vagia, Transeth & Fjerdings, 2016). However, these advances are not limited to the optimization of automation processes, and, therefore, not relegated to the simple delegation of functions and operations to technical artefacts. Rather, through the rise of artificial intelligence and machine learning, technology has obtained abilities that formerly were reserved solely for humans (Jordan & Mitchell, 2015). Because of these two developments (that can also be referred to as automation and autonomization), human–machine collaboration has not only become more prevalent, but also more complex.

For describing human–machine interaction, the concept of socio-technical systems has been introduced. The term indicates that technical components and people are included as inherent parts of a system (Sommerville, 2007). In its simplest form, a socio-technical system is composed of one human and one technical component. Of course, sociotechnical constellations usually take on more complex forms involving many different technical agents and different people or social systems (e.g., organizations). However, every analysis of a socio-technical system implies the analysis of a human–machine collaboration which fulfils a certain function for a certain duration, regardless of its complexity. As the implementation of technology has social, ethical, legal, and economic implications, the setup of socio-technical systems calls for careful consideration (Martin, 2018), requiring one to define the distribution of roles according to ability, authority, and control (Flemisch et al., 2012). The design of

human-machine collaboration affects its governance (Danaher et al., 2017) and accountability (Shin & Park, 2019), as both go hand-in-hand with the underlying distribution of agency (Beck, 2015; Matthias, 2004; Pagallo, 2017). Thus, it shapes the accompanying monitoring measures, as well as the system's administration (Shin & Park, 2019).

Due to the importance of the constitution of socio-technical systems and their underlying human-machine collaboration, these phenomena pose a challenge for management, the law and other disciplines. Organizations and managers must oversee possible changes that the implementation of a new technical system might bring (Shneiderman 2016) and understand its broader impacts (Nunes & Jannach, 2017; Shin & Park, 2019). However, decision makers are often unfamiliar with the fundamentals of implementing human-machine collaboration (Shin & Park, 2019). Managing technology requires a full grasp of the nature of the human-machine collaboration. Comparably, legal assessment equally asks for capturing the distribution of agency, because it significantly affects the attribution of responsibility (Simmler, 2019). The digital age confronts the law with a reality in which human-computer interaction is the rule rather than the exception (Wein, 1992), challenging practitioners to deal with technology with which they are unfamiliar.

Understanding the core of a socio-technical collaboration is no longer a task solely for engineers, but also for decision makers and experts in different fields and disciplines. As illustrated above, not only answering questions regarding if, but also how to implement human-machine interaction within socio-technical systems are decisive managerial tasks. Above all, such implementation requires that those responsible understand and capture the essence of human-machine collaboration. Therefore, the goal of this research is to provide a comprehensive yet straightforward taxonomy for classifying human-machine collaboration. Taxonomies are valuable tools for understanding and analyzing complex phenomena (Nickerson, Muntermann, Varshney & Isaac, 2009). Determined by their aim (Lambe, 2007; Nickerson et al., 2009), they store knowledge in a concise form and reduce complexity

(Lambe, 2007). The goal of the taxonomy presented here is to categorize human-machine collaborations by their most fundamental socially perceived characteristics. Decision makers and researchers will now be equipped with a tool that enables them to quickly grasp the implications of the human-machine collaboration at issue. A taxonomy that focuses on the social perception of socio-technical phenomena will help to evaluate such collaborations in terms of their consequences for different areas and with respect to a diversity of aspects.

Following this objective, this research first describes the method by which this taxonomy was developed. Next, existing taxonomies and classifications are introduced, leading to the elaboration of the present taxonomy's theoretical foundation. Hereafter, the taxonomy is developed and described in detail. In addition, the approach is further clarified with examples. Finally, the scope and limitations of the taxonomy and avenues for further research are discussed before conclusions are drawn.

2 Methodical framework

The approach to developing this taxonomy consisted of a literature review and consideration and elaboration of existing taxonomies and categorizations. We searched the most common databases using terms such as “technology,” “classification,” and “taxonomy.” Whenever possible, we based our analysis on existing taxonomies and classification systems. However, because some of these did not fully meet our objectives, we extended and adapted what we found. Based on the results of this process, we followed three steps to develop the taxonomy presented here.

We first evaluated the core parameters and axes. In so doing, we evaluated whether each parameter captured the essence of human-machine interaction. A parameter was considered relevant if it significantly shaped the nature of the human-machine collaboration and clearly displayed differences between the human and technical components.

In the second step, we derived the parameter levels. We mapped each level on a continuum ranging from “none” to “all.” We introduced new levels when there were significant shifts in social and technical differences (i.e., when a technical component was assigned a new task or capability). To determine these levels, we focused especially on their (social) implications (i.e., their impact on social perceptions and the attribution of responsibility) and their influence on accountability.

In a third step, we validated the parameters and levels with regards to their clarity, comprehensibility, and practicability. In this step, we wanted to make sure that the taxonomy would be useful to people with little or no technical expertise. The goal was to present an instrument for capturing the basic characteristics of interdisciplinary relevance of human–machine collaboration.

3 Theoretical Background

3.1 Conceptualizing socio-technical phenomena

In the past several decades, there has been a multitude of approaches to capturing human–machine collaboration. The long-standing tradition has focused on different levels of automation (Vagia et al., 2016). These taxonomies describe different role allocations (Kaber, 2018) and offer descriptions of what tasks the “human operator” and “computer” are assigned within a collaboration. Usually, the levels of automation range from fully manual to fully automated (Vagia et al., 2016). Fully manual describes a situation in which the human is fully in charge. Conversely, fully automated indicates that the human operator is completely out of the loop and, therefore, is obsolete (Parasuraman, Sheridan & Wickens, 2000). Such taxonomies center on the interaction between human and machine, the level of which is derived from variations in task allocation (Endsley, 1987; Sheridan & Verplank, 1978; Weyer, 2006). However, there seems to be less consensus about what and how many levels lie in between these points. Some approaches describe four (Endsley, 1987), others ten

(Sheridan and Verplank, 1978), and some even more (Riley, 1989). A recent study described different levels of automation with regards to autonomous driving (NHTSA, 2013). There, the scale ranged from “no automation” to “autonomous”, implying that autonomy is the highest form of automation and reflecting a slightly different stance. Solely focusing on role allocation, however, is not the only way of describing human–machine interactions. There are models that also take into account the process by which a task is completed. By categorizing every stage of a given process or decision, the level of automation can be distinguished even more precisely, depending on what stage of the process is being automated and to what extent (Parasuraman et al., 2000; Proud, Hart & Mrozinski, 2003; Weyer, 2006).

A different research stream centers on a technology’s abilities. One line of thought has focused on naming and discussing the capabilities of an advanced technology, without further mapping. In this debate, scholars have discussed conditions for technical autonomy and agency. The capacity to react to changes in the environment (Alonso & Mondragón, 2004; Franklin & Graesser, 1997) and the general ability to interact (Floridi & Sanders, 2004; Misselhorn, 2015) have both been proposed. Adaptability has been repeatedly identified as relevant, proposed as either an optional (Franklin & Graesser, 1997) or necessary (Floridi & Sanders, 2004; Misselhorn, 2015; Russell & Norvig, 2014) condition for technical agency. An autonomous agent has been described as having the ability to perceive its environment and then learn from and adapt to it (Alonso & Mondragón, 2004; Pagallo, 2017; Santosuosso & Bottalico, 2017; Sartor & Omicini, 2016) or to sense, plan and intentionally act upon an environment without external control (Beer, Fisk & Rogers, 2014). A study investigating what consumers regard as “intelligent” when it comes to technology came to similar conclusions. Product intelligence is said to be composed of the key dimensions of autonomy: an ability to learn, reactivity, an ability to cooperate, humanlike interaction, and personality (Rijsdijk, Hultink & Diamantopoulos, 2007).

While these approaches suggest preconditions and basic characteristics, none conceptualize connectivity among the parameters or clearly topologize their

impact on autonomy. Yet with regards to technical agency, there have been a number of approaches suggesting different stages or levels. According to these approaches, technical agency is not a binary category, but rather varies in terms of degree. For example, it was suggested that one must distinguish different activity levels when dealing with technical agency, using a scale ranging from “passive” (describing artefacts operating as mere tools) to “transactive” (indicating actions based on intelligent associations stemming from self- and external reflection) (Rammert, 2009). Advanced technology is assumed to challenge the passive nature of technology, blurring the boundary between technical functionality and human agency. Rammert and Schulz-Schaefer (2002) identified three levels of agency: causality as the capacity to cause change, contingency as the ability to recognize action alternatives, and the highest level, intentionality, which is defined as intentional reasoning engaging in action. Thürmel (2015) added another gradualization by proposing that the degree of agency is constituted by four dimensions: activity and adaptivity for individual agency as well as interaction and personification of others for joint agency. There have also been other approaches to combine different variables into one categorization of technical systems, such as autonomy, field of application, and morphology (Onnasch, Maier & Jürgensohn, 2016) or automation and complexity (Janssen & Kuk, 2016), the latter following the assumption that the more complex a technical system is the opaquer it gets. While these approaches mainly center on the technical side, the typology provided by Gransche, Shala, Hubig, Alpsancar and Harrach (2014) additionally describes the human role in human-technology interaction, while comparing and linking technical and human autonomy with control. According to the authors, the “normative” type of autonomy and control is reserved only for humans; technical systems achieve “strategic” and “operative” autonomy and control (Gransche et al., 2014).

3.2 Synthesis

To sum up, there are different taxonomies for describing human–machine collaboration, and these follow different aims. However, the literature review has shown that there are two basic questions underlying these existing frameworks. One stream describes different role allocations between humans and technical systems and asks who does what? (Kaber, 2018). This is the core question supporting the concept of automation. Taxonomies centering on automation usually address information systems experts and their pursuit of technical optimization. Such taxonomies are accurate in their capture of basic design issues. Nonetheless, task allocation is not the only point to consider when implementing and optimizing human–machine collaboration. The second stream takes technical functions and properties into account more directly. They focus on intelligence (Rijsdijk et al., 2007), autonomy (Beer et al., 2014), or agency (Thürmel, 2015), and, therefore, on the technical system’s abilities. These approaches are strongly interwoven, and different scholars have reached similar conclusions indicating that the primary concern is to describe a technology’s independence. The basic question behind these approaches is: what degree of freedom does the technology have when completing the assigned task? Therefore, they center on technical autonomy.

In sum, a first analysis revealed that the underlying most fundamental characteristics of human–machine collaboration are automation and autonomy. Current approaches to understanding and conceptualizing such socio-technical interaction focus either on automation or technical autonomy (or comparable parameters). Sometimes, these notions are used interchangeably (Vagia et al., 2016) or mapped onto the same continuum (NHTSA, 2013). We suggest that the allocation of situational executive control (automation) and independence of the technical component assigned (autonomy) only in combination comprise the core parameters of human–machine collaboration.

4 Taxonomy

4.1 The axes: automation and autonomy

Although there is no universal definition, it is generally agreed that automation refers to a state in which a technical system performs a formerly human task or parts of that task, respectively (Parasuraman et al., 2000; Vagia et al., 2016). In fulfilling this task, the machine runs without a human operator (Hertzberg, 2015; Nof, 2009). As explained above, the task assigned to the technical system might vary in its scope, i.e., be partially or fully automated (Vagia et al., 2016). With regards to human-machine-collaboration, the level of automation describes the extent of the machine's contribution to the joint performance. It is crucial to determine whether and to what extent humans are still "in the loop" with regards to acting and decision-making (Santosuosso & Bottalico, 2017; Sartor & Omicini, 2016). Designing human-machine collaboration requires the careful determination of the tasks to be automated, and to what extent (Parasuraman et al., 2000). Thoughtful design decisions depend upon a thorough understanding of such task allocation (Vagia et al., 2016).

While an appropriate definition of the level of automation is crucial for capturing this allocation of control, it neglects the nature of advanced technologies. Advanced machines can be more or less independent from their human creators and operators (Rammert & Schulz-Schaefer, 2002). Technical autonomy addresses this notion. Traditionally, autonomy is understood as self-governance, self-sufficiency, or self-directedness (Bradshaw et al., 2004; Vagia et al., 2016); it is a relational concept describing the degree of independence from something, such as the specific influence of another entity, the environment, or internal restraints (Castelfranchi & Falcone, 2004; Müller-Hengstenberg & Kirn, 2016; Verhagen, 2004). Usually, the possibility of independently developing action alternatives and acting according to one's own preferences serve as prerequisites for autonomous behavior (e.g., Müller-Hengstenberg & Kirn, 2016). In a Kantian view, autonomy requires the ability

to adopt maxims to govern one's action and involves the power to make laws for oneself (Hilgendorf, 2017; Korsgaard, 2014; Misselhorn, 2015). This association with Kant's moral philosophy requires a careful use of the term; however, its use in describing technical agents is far less demanding (Misselhorn, 2015). As mentioned above, the debate surrounding the autonomy of advanced technology is closely related to the discussion of the possibility of and conditions surrounding technical agency per se, as these are often regarded as mutually conditional (Misselhorn, 2015). Autonomy is, therefore, closely linked to the concept of agency.

How autonomously a technical component operates is important to understand, because this affects not only the limits of human control, but, furthermore, also other variables such as the explainability, traceability and, predictability of the technical component's action (Balkin, 2015). Thus, for an initial assessment, especially in terms of consequences, it is important to determine how the task is fulfilled by the technical system. Therefore, looking at technical autonomy is equally pivotal. Evaluating a human-machine collaboration demands the determination of both values: the extent to which the task is automated (automation) and how autonomous the technical system operates when pursuing that automated task (autonomy). We conceptualize both automation and autonomy as gradual, indicating that these are not binary categories but rather vary in degree. This suggests a taxonomy composed of two axes that consist of different levels.

This taxonomy focuses solely on technical autonomy. Conversely, the autonomy of the human as a part of this human-machine collaboration is generally regarded as given and more substantial than what technology can ever reach (Misselhorn, 2015; Rammert, 2009). Deliberately focusing on technical autonomy, this taxonomy does not map or describe the concept of human autonomy in detail. A further division of human autonomy would not add value to the understanding of human-machine collaboration. Although the taxonomy focuses on technical autonomy, its description still includes a human perspective: the extent to which the actions of technology are comprehensible

to human counterparts. Thus, the levels of technical autonomy include the human perspective within the human–machine collaboration, though not directly.

The following examples highlight the difference between automation and autonomy and clarify the benefit of including these two axes in one taxonomy. A physician and diagnostic program together collaborate when they pursue the task of diagnosing patients. Fulfillment of this task can be automated to different degrees (i.e., a computer program can be assigned a varying part of the task). It can, for example, suggest diagnoses when certain symptoms are entered or the program can reach a single diagnosis and even order medication without human influence. In addition, the technical component can be more or less autonomous in a sociotechnical constellation. For example, the diagnostic system can only be a very cognitively limited computer program with a restricted dataset and clearly defined variables and algorithms. If it is more autonomous, it is an open system that obtains data from networks or other technical agents. Furthermore, it can learn from diagnoses already conducted and adapt its behavior accordingly. These differences relate to its level of autonomy.

A self-driving car will serve as a second example. Here, the task of operating the car is fully automated. No human intervention is needed. Therefore, a self-driving car reaches the highest stage of automation. In terms of autonomy, however, the evaluation may vary from car to car. There are self-driving cars whose operation is fully determined, and a given body of information always leads to the same predefined output. Due to the absence of machine learning, the car does not alter its behavior. Yet there may also be cars that reach notably higher levels of autonomy. Some self-driving cars not only learn and adapt, but are also connected to other agents, resulting in an open, multi-agent system. Consequently, even within the scope of fully automated driving, there is a range of different types of human–machine interplay, as the autonomy of the technical aspect can vary significantly.

To again underscore the important distinction between the two axes and the value they add to accurately capturing socio-technical systems, it is useful to revisit what lies at the taxonomy's core. It centers on human-machine collaboration of a certain permanence and distinctiveness, consisting of one or multiple human actors and one or multiple technical agents (Weyer & Reineke, 2005). If one focuses only on a single task taken over by a machine, that task can be described as embodying full automation. However, conceptualizing socio-technical phenomena means focusing on a human-machine collaboration not only executing a single act, but also fulfilling a distinct and permanent function. During this persisting collaboration, the system as a whole can be more or less automated, and within these automated tasks, the machine can be more or less autonomous. Both axes define the human-machine interaction at its core.

4.2 Levels of automation

We draw our approach to defining the levels of automation from Endsley's (1987) four levels, because this study is concise and comprehensible and does not neglect the core aspects of task allocation in the context of human-machine collaboration. We complemented Endsley's levels of automation with an additional level borrowed from Weyer (2006). Different from most taxonomies (Vagia et al., 2016), we did not include a level for manual (and, therefore, solely human) operations, because our taxonomy was designed only to classify human-machine collaboration. This implies that there is already a minimum amount of technical participation, which renders a level for classifying manual tasks unnecessary. The five stages are shown in Table 4 and described in detail thereafter.

Table 4: Levels of automation with their main features.

Level	Description	Explanation
1	Offers decision	Technical component suggests options and the human decides.
2	Executes with human approval	Technical component acts after human approves.
3	Executes if no human vetoes	Technical component acts unless human vetoes.
4	Executes and then informs	Technical component acts independently and human is informed about the actions carried out.
5	Executes fully automated	Technical component carries out actions independently without informing human.

4.2.1 Level 1: Offers decision

Our taxonomy begins with what can be called decision-support systems (i.e., constellations in which the technical component makes suggestions). The machine, usually a computer program, makes recommendations to the operator. The operator chooses whether to act accordingly (Endsley, 1987). In this early stage of automation, the system presents options; however, it is still the human operator who selects and, thus, decides.

4.2.2 Level 2: Executes with human approval

In the second level of automation, the machine executes with human approval. Here, the technical component makes a recommendation that it carries out if the operator concurs (Endsley, 1987). This marks an increase in automation in that the system selects among options and presents only the “best” alternative. The selection process is no longer carried out by the human operator, but rather by the technical component. However, at this level it is still the human who makes the final decision, either by approving or rejecting the system’s proposition.

4.2.3 Level 3: Executes if no human vetoes

In our third level of automation, the technical system executes if no human veto occurs. According to Endsley (1987), this means that the system recommends a single option, which it will carry out unless the operator vetoes that decision. This represents an increase in automation, because the technical system selects an option and automatically decides. However, the human operator remains a part of the process, because the operator can correct the action through its veto power. The decisions made by the system still require human consent. However, this no longer happens in advance, only after the component has already reached a decision. This influences the social perception of the operator's role.

4.2.4 Level 4: Executes and then informs

Further increasing the automation, at this level the technical system executes and then informs. We borrowed this level from Weyer (2006) and use it to augment Endsley's (1987) approach. The system selects an option and then automatically acts accordingly. However, the human operator still has a role in the process, because the system informs the human when it acts. Even if a human operator no longer plays an active role, they remain informed about the system's actions. The knowledge of these actions is relevant for social perception. Therefore, it differs from Level 5, where no human is made aware of the automated process.

4.2.5 Level 5: Executes fully automated

At the fifth level, the technical system executes fully automated. This means that the human is neither informed about nor actively participates in the action. The technical system is fully in charge, rendering human action obsolete.

4.3 Levels of autonomy

There was no classification of technical autonomy available that met with the scope of our taxonomy (i.e., that described the independence of the technical component while completing an assigned task). Therefore, we could not simply borrow different levels of autonomy as we did when typologizing automation. After reviewing the literature, we isolated the four dimensions we considered most crucial when determining the different levels of autonomy.

We tend to call something “autonomous” if it appears untraceable. Technical autonomy is generally said to describe systems that to a certain degree are independent and not fully determined (Verhagen, 2004; Castelfranchi & Falcone, 2004; Müller-Hengstenberg & Kirn, 2016). Thus, technical autonomy involves a certain lack of traceability and certainty regarding the system’s performance, and concerns both, transparency and determinism. Conversely, the absence of technical autonomy means full determination and traceability. A technical system that meets this condition is called deterministic. In a deterministic system, input A always equally leads to output B, and all execution steps in between are specified and transparent (Loh & Loh, 2017). A technical system is considered determined when every possible condition of the system unambiguously results in a subsequent condition of $n+1$ (Müller-Hengstenberg & Kirn, 2016). However, such a determined system does not necessarily need to be transparent in all of its execution steps, leaving the observer or user ignorant about the detailed means of processing. Put differently, a deterministic system is fully transparent and traceable; a non-transparent system is not because of the opacity of the steps required to reach a specified output. Transparency in a technical system allows for easier tracking and reconstruction, which gains relevance in questions of responsibility (Mittelstadt, Allo, Taddeo, Wachter & Floridi, 2016) and influences social perception. Consequently, regarding certainty of output and traceability as crucial for human control consequently led us to the first and second dimensions of technical autonomy: non-transparency and indetermination.

As introduced above, the literature review revealed that adaptability is usually considered crucial for technical autonomy. We follow this position and consider adaptability to be the third core dimension. Being autonomous requires the capacity to learn and adapt behavior to a changing environment (Alonso & Mondragón, 2004; Floridi & Sanders, 2004; Pagallo, 2017; Santosuosso & Bottalico, 2017; Sartor & Omicini, 2016; Thürmel, 2015). A machine of this kind is able to process information, expand the knowledge implemented by programmers, and change the way it responds (Müller-Hengstenberg & Kirn, 2016; Sartor & Omicini, 2016; Thürmel, 2015). This allows the system to adapt and to improve its performance in a certain environment (Sartor & Omicini, 2016) without human intervention (Floridi & Sanders, 2004; Pagallo, 2017). Thus, adaptable systems are capable of altering their behavior which tends to make them more unpredictable for and independent from human operators. Adaptability, therefore, shapes technical autonomy.

Interactivity is also regarded as a crucial feature of autonomy, as was revealed in the above discussion of core parameters. Advanced forms of interactivity result in system openness, what we consider the fourth key dimension when determining the level of autonomy. Various authors have recognized expansion of the original data through collaboration with other agents as a characteristic of autonomy and called this ability cooperation (Müller-Hengstenberg & Kirn, 2016) or interactivity (Floridi & Sanders, 2004; Pagallo, 2017; Sartor & Omicini, 2016). Due to collaboration, autonomous systems can delegate and divide labor among on another or build multi-agent systems (Müller-Hengstenberg & Kirn, 2016). The shared ability to solve problems in a multi-agent system can exceed the capability of a single agent (Thürmel, 2015; Weyer & Reineke, 2005). The “intelligence” of multi-agent systems emerges through coordination among agents (Rammert & Schulz-Schaefer, 2002; Weyer & Reineke, 2005) and hence is not predefined by programmers (Müller-Hengstenberg & Kirn, 2016). It is not foreseeable with whom these technical systems will interact (Alonso & Mondragón, 2004) and where they will gather their data, e.g., because the technical system is connected to other software agents. Openness with respect

to such cooperation results in greater technical autonomy as does the openness and flexibility of the system with regards to data-gathering within the environment (and in other ways such as through the internet). Technical systems that are open in the sense that input is neither predefined nor limited are increasingly unforeseeable, resulting in the social ascription of more autonomy.

To sum up, due to a technical system's autonomy it is unforeseeable with whom it will interact and from where it will gather its data (openness). Furthermore, it can alter its behavior on the grounds of experience (adaptability). Its programming can leave it undetermined (indetermination) and untraceable (non-transparency). Our conceptualization thus identifies four dimensions defining the level of technical autonomy: non-transparency, indetermination, adaptability, and openness. Hence, to determine how autonomous a technical component is, one must answer the questions listed in Table 5.

Table 5: Dimensions of technical autonomy with assessment questions.

Dimension	Assessment Question
Non-Transparency	Is the user and/or observer able to trace how the system gets from Input A to Output B?
Indetermination	Does Input A always lead to the same Output B?
Adaptability	Is the same system able to learn from its experience?
Openness	Is the system able to expand its original input in particular due to cooperation with other systems and/or its flexibility in gathering source data?

This binary distinction is a simplification, and all of these characteristics can be present to varying degrees. However, this simplification is necessary to reduce complexity and make the taxonomy practical. Once the crucial features of the technical component are defined and the presence and absence of each dimension evaluated, the level of autonomy can be determined. The definitions of the taxonomy's levels of autonomy rely on the particular combination of features, as presented below and illustrated in Table 6.

Table 6: Levels of autonomy with an explanation of the main features.

Level	Name	Description
1	Deterministic System	Technical system is determined, transparent, unadaptable, and closed.
2	Non-transparent System	Technical system is determined, <i>non-transparent</i> , unadaptable, and closed.
3	Indetermined System	Technical system is <i>indetermined</i> , non-transparent, unadaptable, and closed.
4	Adaptable System	Technical system is indetermined, non-transparent, <i>adaptable</i> , and closed.
5	Open System	Technical system is indetermined, non-transparent, adaptable, and <i>open</i> .

4.3.1 Level 1: Deterministic system

In the first and lowest level of autonomy, the technical component is transparent and determined, meaning that a given input always leads to a specified output, with full transparency with regards to how the system reaches that output. In such a machine, everything is completely predefined; the system is closed and has no ability to learn. The system is fully traceable and predictable. Simple calculators for example would operate on this level.

4.3.2 Level 2: Non-transparent system

A non-transparent technical system operates at the second level of autonomy. This means that the system remains fully determined but not every step is predefined and traceable. This marks an increase in autonomy, because the component potentially alters its manner of moving from the input to the output, and thus is non-transparent in its procedures. It holds back information and becomes opaque to the human operator or observer, which impacts social perception. However, such a system's output is still determined, because a certain input always leads to a particular output. An example for this level is a system that weights different parameters when reaching a decision. While the

same input variables always lead to the same result, the human operator cannot fully trace how the relevant parameters are weighted in each single case.

4.3.3 Level 3: Indetermined system

At the third level of autonomy, the technical component is indetermined and non-transparent. How a given input leads to a particular output is untraceable. The component is even more opaque, because it does not provide complete information about the execution steps and alters the output. At this level, there is no certainty of the result (i.e., the output is not determined). The same input does not necessarily result in a particular output; the system is not predictable. A common Level 3 application is a chance factor that is purposefully implemented to vary the output. Likewise, programming that links the system to a specific (yet at the moment of programming unknown) external input variable is a common characteristic of such applications.

4.3.4 Level 4: Adaptable system

At level four, the technical system is able to learn from its own data. Once a system can learn, it can no longer be completely determined or tracked, because the input and pathway for reaching the output may change. Due to machine learning, an intelligent system possesses a considerable amount of autonomy. Such a system is not only undetermined due to a specifically implemented and randomized variable, it is capable of changing its behavior base and becomes adaptable to the environment. The output and single steps toward execution are thus variable and can be permanently adapted. This again marks an increase in opacity as it hampers the comprehensibility of the system's action. The human part can no longer comprehend the criteria and circumstances by which the actions of the technical system are coordinated. A technical system drawing upon a machine learning algorithm would be an example for this level of autonomy.

4.3.5 Level 5: Open system

A system is classified as Level 5 if it is undetermined and non-transparent. The former implies that a given input must not lead to the same output, while the latter indicates that the way it reaches the output is not easily accessible by the observer. Furthermore, it is adaptable due to its ability to learn, leaving room for modifications in the way it completes a task. It is capable of interacting with other systems and able to collect data. This openness marks the highest level of autonomy, because the system is neither limited to a specified and predefined input, nor to its own experience. Data gathering for the system's input is not under the full control of a human operator or programmer. It is, therefore, even more difficult to trace a system's actions, because those actions are opaque. A technical system based on a machine learning algorithm and additionally connected to an internet database, where it has access to new learning data would be an example for this fifth level of technical autonomy.

5 Discussion

The goal of this article was to introduce a tool that would allow everyone, regardless of technical expertise, to capture the basic structure of any human-machine collaboration. Two parameters lie at the core of our taxonomy that essentially characterize human-machine collaboration: automation and autonomy. We propose that every human-machine collaboration is typified mainly and most fundamentally by the nature of this cooperation (i.e., the distribution of tasks between the human and technical component) and the technical components' abilities and capacities, in other words, by the levels of automation and technical autonomy. Figure 1 maps these two axes and depicts the proposed taxonomy.

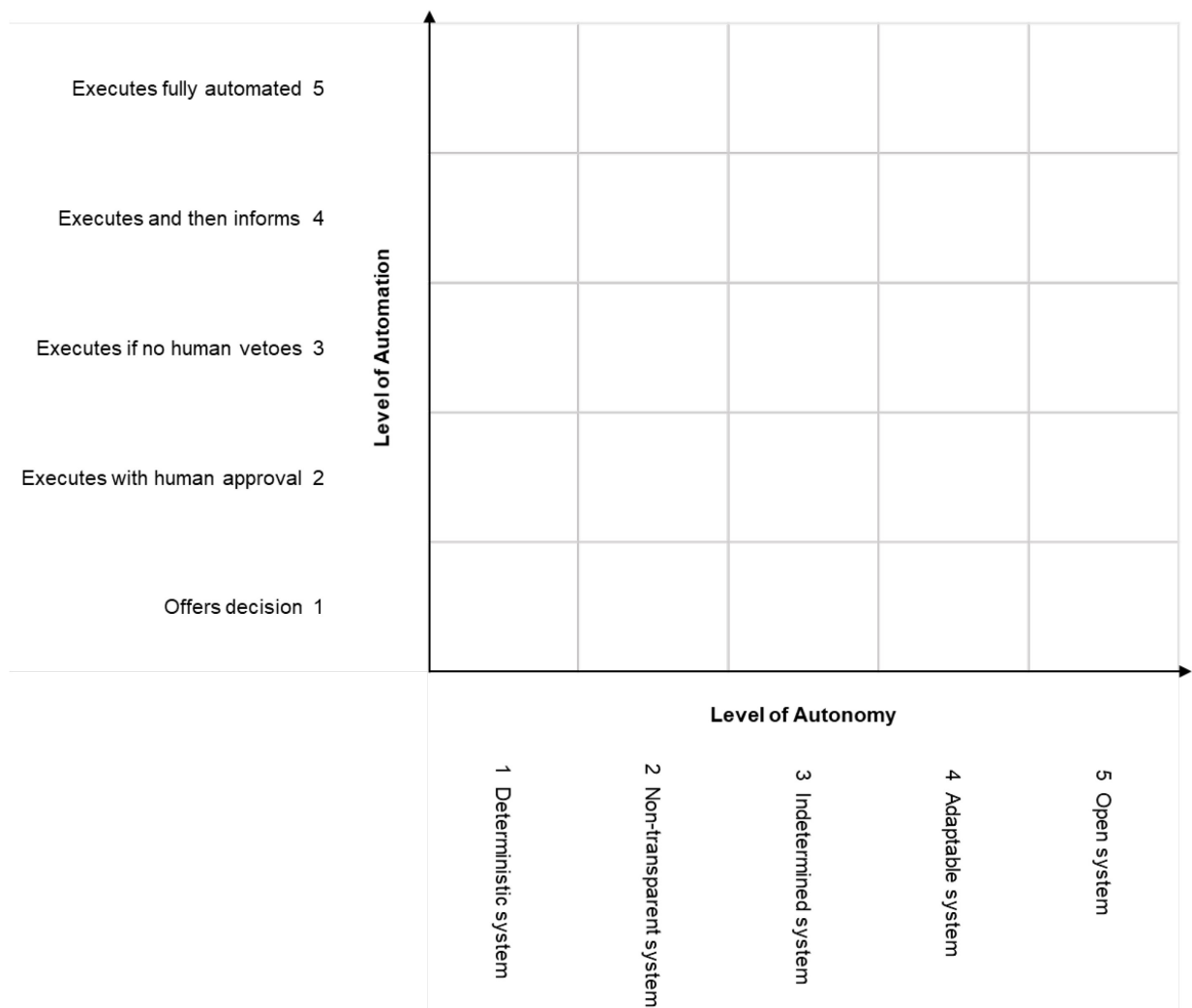


Figure 1: Taxonomy of automation and technical autonomy in human-machine collaboration.

This taxonomy unites two research streams that run parallel to one another: the human factor and automation research (e.g., Endsley, 1987; Parasuraman et al., 2000; Vagia et al., 2016), and technical sociology (Rammert & Schulz-Schaefer, 2002; Thürmel, 2015) and computer ethics (Floridi & Sanders, 2004) and thus rests on the concept of technical autonomy. Approaches that focus either on automation or on technical autonomy are not sufficiently differentiated. Only the combination of the two dimensions allows for a thorough understanding of human-machine collaboration, as both pose fundamental challenges. Understanding the level of automation is important, because the consequences are potentially far reaching; the allocation of control, among other aspects, affects the distribution of agency and, thus, responsibility (Chinen, 2016; Kirchkamp & Strobel, 2019). The higher the level of automation, the less human control there is over the system's actions. Thus, increasing levels of automation ask for higher requirements for example regarding precaution measures, otherwise potentially resulting in liabilities. Conversely, discerning the level of autonomy is crucial because this affects traceability and comprehensibility (Zarsky, 2016). The greater the level of autonomy, the opaquer and more intractable the machine's actions appear. This necessitates greater requirements in terms of comprehensibility and monitoring, as the human part of the system should still be able to understand and explain what the technical part has done. Challenges regarding a lack of explainability arise, likewise affecting questions of legitimacy (Mittelstadt et al., 2016), governance (Danaher et al., 2017) and accountability (Shin & Park, 2019).

What a taxonomy does or does not capture is a chosen and, therefore, variable specification. Depending on the perspective, this specification may change. Automation and autonomy are not the only possible dimensions for describing advanced technology. There are, for example, taxonomies that classify socio-technical systems according to their degree of automation and complexity (Janssen & Kuk, 2016). However, complexity is a consequence of advanced technology, and, therefore, not necessarily a core element of human-machine collaboration. Other scholars have concerned themselves with (technical)

agency per se (Rammert, 2009; Thürmel, 2015). Although their analyses contribute very important and fundamental aspects that significantly influence this work, these studies have not allowed for a schematic assessment of the division of agency in terms of the concrete components determining the distribution itself. Other conceptualizations match different types of autonomy with different types of control in human-technology interaction, concluding that only humans exert the highest levels (Gransche et al., 2014). This assignment of different forms of autonomy and control to humans and technology is useful. It does not, however, explain in detail what characteristics and capabilities shape technical autonomy. Furthermore, as has been referred to throughout this work, there are many taxonomies and schemes that center solely on automation (see, e.g., Vagia et al., 2016). These neglect a dimension crucial to social perception, and, therefore, to many implementations.

Both, the dimensions, and number of levels, depend on the perspective and particular taxonomy's aims. For example, scholars have reached different conclusions about the number of levels of automation (Vagia et al., 2016) or of autonomy (Gransche et al., 2014) necessary for adequate differentiation. Defining the appropriate number of levels is not, however, the most fundamental concern, as that number reflects only precision and usability. What matters more is that automation and autonomy are understood as gradual. While this has long been the case for automation (e.g., Endsley, 1987; Sheridan & Verplank, 1978), the understanding of technical autonomy as gradual is still relatively new (cf. for agency see Rammert, 2009). Understanding the graduality of both dimensions contributes to a more differentiated capturing of advanced technology. Furthermore, it helps to more precisely highlight differences among human-machine collaborations.

5.1 Limitations

This taxonomy only allows for the classification of sociotechnical collaboration within the boundaries of functional fulfillment. When evaluating such systems, the levels of automation and technical components' autonomy must be Fig. 1 Taxonomy of automation and technical autonomy in human-machine collaboration specified. However, it is equally important to determine the function that a particular system fulfils (i.e., what task it is assigned). Appraising the legitimacy of the technical system requires careful consideration of whether the task is, in the first place, suitable for automation and autonomization. However, this feature cannot be captured abstractly. Therefore, this consideration cannot be part of the general taxonomy, but rather serve to complement an assessment. This taxonomy also captures technical autonomy only in terms of what is possible today. It might well be that further technological developments lead to new and more advanced technological systems. However, if this is the case, our taxonomy can easily be extended and refined. It is important to note that this study's philosophical touchstones are not a limitation. In developing this taxonomy, concepts with substantial philosophical loading were unavoidable. Autonomy, as well as the states of being determined (and determinism) and undetermined (and indeterminism) are often used in technical literature. Employing them here should not lead to the misunderstanding that technology can assume these demanding philosophical requirements in the same ways as humans.

5.2 Avenues for further research

This taxonomy offers a number of initial points for future research. From a legal science perspective, investigating the influence of different levels of automation and autonomy on the attribution of legal responsibility is an interesting endeavor. Likewise, empirically testing if and to what effect the different levels have on punishment would be a fruitful field of research. From a management perspective, future work should derive implementation standards and

precaution measures from this taxonomy, as well as describe requirements for organizations to use when implementing socio-technical collaboration. One example is a division into three levels, distinguishing between low-, medium-, and high-level implementation of human-machine collaboration. This categorization follows the assumption that a low level of autonomy or automation would lead to a low-level implementation, which in turn would require less monitoring and fewer precautions. Conversely, high values on both axes would indicate a high-level implementation, which would potentially be more demanding. Future research could also draw attention away from the collective and towards the individual level. Scholars could explore whether these levels have an actual effect on social and psychological perception, and if so, how this perception changes if automation and autonomy increase. One could also investigate what influence these levels have on decision-making and how they affect trust. In sum, there is a wide range of psychological, sociological, legal, and managerial research questions that could be explored based on this taxonomy, which will serve as a tool for clearly differentiating among the varieties of distributed agency in human-machine collaboration. Moreover, these findings will have a direct impact on policy discussions accompanying the implementation of technology in the digital age.

6 Conclusion

As was stated at the beginning of this work, the spread of advanced technology has resulted in human-machine collaboration becoming more and more prevalent. The consequences of their spread are far-reaching. Describing and understanding human-machine interaction is, therefore, crucial and should no longer be a task for engineers, software developers, or systems designers; it is now important for professionals and researchers in a wider variety of fields. However, socio-technical constellations are complex, and capturing them adequately calls for a distillation to their basic characteristics involved in human-machine collaboration: automation and autonomy. Introducing this

taxonomy based on these two dimensions may at first glance seem bold. It is important to note, though, that this boldness adds value. It allows the user to grasp complex phenomena and focus on what is most important in human-machine collaboration: the question of who does what on one hand, and the question of how independent is it done on the other hand. This clear description of the distribution of agency will allow professionals and researchers from different fields to estimate and evaluate the implications and consequences of given socio-technical constellations. We do not seek to oversimplify the discussion of socio-technical collaboration. To the contrary, by introducing different levels of automation and autonomy, we emphasize that these concepts are not a question of all or none, but rather vary in degree. The gradual approach presented here allows for a more differentiated description and understanding. Thus, the taxonomy points out that human-machine collaboration is multifaceted, while also allowing for all users, despite their level of specialized knowledge, a pragmatic means of capturing them.

7 References

- Alonso, E. & Mondragón, E. (2004). Agency, learning and animal-based reinforcement learning. In M. Nickles, M. Rovatsos M & G. Weiss (Eds.), *Agents and computational autonomy – potential risks and solutions* (pp. 1-6). Berlin/Heidelberg, Springer.
- Balkin, J. M. (2015). The path of robotics law. *California Law Review*, 6, 45.
- Beck, S. (2015). Technisierung des Menschen – Vermenschlichung der Technik. Neue Herausforderungen für das rechtliche Konzept „Verantwortung“. In: M. C. Gruber, J. Bung, & S. Ziemann (Eds.), *Autonome Automaten: Künstliche Körper und artifizielle Agenten in der technisierten Gesellschaft* (pp 173 – 187). Berlin: BWV Verlag.
- Beer, J. M., Fisk, A. D., & Rogers, W. A. (2014). Toward a framework for levels of robot autonomy in human-robot interaction. *Journal of Human-Robot Interaction*, 3, 74-99.
- Bradshaw, J. M., Feltovich, P. J., Jung, H., Kulkarni, S., Taysom, W., & Uszok A. (2004). Dimensions of adjustable autonomy and mixed-initiative interaction. In M. Nickles, M. Rovatos, & G. Weiss (Eds.). *Agents and computational autonomy – potential, risks, and solutions* (pp. 17-39). Berlin/Heidelberg: Springer.
- Castelfranchi, C. & Falcone, R. (2004). Founding autonomy: The dialectics between (social) environment and agent’s architecture and powers. In M. Nickles, M. Rovatos, & G. Weiss (Eds.). *Agents and computational autonomy – potential, risks, and solutions* (pp. 40-54). Berlin/Heidelberg: Springer.
- Chinen, M. A. (2016). The Co-evolution of autonomous machines and legal responsibility. *Virginia Journal of Law & Technology*, 20, 338-393.
- Danaher, J., Hogan, M. J., Noone, C., Kennedy, R., Behan, A., De Paor, A., . . . Shankar, K. (2017). Algorithmic governance: Developing a research agenda through the power of collective intelligence. *Big Data & Society*, 4(1), 1-21. Doi:10.1177/2053951717726554
- Endsley, M. R. (1987). The application of human factors to the development of expert systems for advanced cockpits. *Proceedings of the Human Factors Society Annual Meeting*, 31,1388-1392. Doi:10.1177/154193128703101219

-
- Flemisch, F., Heesen, M., Hesse, T., Kelsch, J., Schieben, A., & Beller, J. (2012). Towards a dynamic balance between humans and automation: Authority, ability, responsibility and control in shared and cooperative control situations. *Cognition, Technology & Work*, 14, 3-18.
Doi:10.1007/s10111-011-0191-6
- Floridi, L., & Sanders, J. W. (2004). On the morality of artificial agents. *Minds and Machines*, 14, 349-379. Doi:10.1023/b:mind.0000035461.63578.9d
- Franklin, S., & Graesser, A. (1997). Is It an agent, or just a program? A taxonomy for autonomous agents. In J. P. Müller, M. J. Wooldridge, N. R. Jennings (Eds.), *Intelligent agents III agent theories, architectures, and languages*. Lecture Notes in Computer Science, (pp. 21-35), Berlin/Heidelberg: Springer.
- Gransche, B., Shala, E., Hubig, C., Alpsancar, S., & Harrach, S. (2014). Wandel von Autonomie und Kontrolle durch neue Mensch-Technik-Interaktionen. Grundsatzfragen autonomieorientierter Mensch-Technik-Verhältnisse. Stuttgart: Fraunhofer Verlag.
- Hertzberg, J. (2015). Technische Gestaltungsoptionen für autonom agierende Komponenten und Systeme. In E. Hilgendorf, & S. Hötitzsch S (Eds.), *Das Recht vor den Herausforderungen der modernen Technik* (pp. 63-74), Baden-Baden: Nomos.
- Hilgendorf, E. (2017). Automated driving and the law. In E. Hilgendorf, & U. Seidel (Eds.), *Robotics, Autonomics, and the Law* (pp. 171-194), Baden-Baden: Nomos.
- Janssen, M., & Kuk, G. (2016), The challenges and limits of big data algorithms in technocratic governance. *Government Information Quarterly*, 33, 371-377. Doi:10.1016/j.giq.2016.08.011
- Jordan, M. I., & Mitchell, T. M (2015). Machine learning: trends, perspectives, and prospects. *Science*, 349, 255-260. Doi:10.1126/science.aaa8415
- Kaber, D. B. (2018). Issues in human–automation interaction modeling: Presumptive aspects of frameworks of types and levels of automation. *Journal of Cognitive Engineering and Decision Making*, 12, 7-24.
Doi:10.1177/1555343417737203
- Kirchkamp, O., & Strobel, C. (2019). Sharing responsibility with a machine. *Journal of Behavioral and Experimental Economics*, 80, 25-33.
Doi:10.1016/j.soccec.2019.02.010

-
- Korsgaard, C. M. (2014). The normative constitution of agency. In M. Vargas & G. Yaffe (Eds.), *Rational and social agency: The philosophy of Michael Bratman* (pp. 190-215). New York: Oxford University Press.
- Lambe, P. (2007). *Organising knowledge: taxonomies, knowledge and organizational effectiveness*. Oxford: Chandos.
- Loh, W., & Loh, J. (2017). Autonomy and responsibility in hybrid systems. In P. Lin, R. Jenkins, & K. Abney (Eds.), *Robot ethics 2.0: From autonomous cars to artificial intelligence*. Oxford: Oxford University Press.
Doi:10.1093/oso/9780190652951.003.0003
- Martin, K. (2018). Ethical implications and accountability of algorithms. *Journal of Business Ethics*, 160, 835-850. Doi:10.1007/s10551-018-3921-3
- Matthias, A. (2004) The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and Information Technology*, 6, 175-183. Doi:10.1007/s10676-004-3422-1
- Misselhorn, C. (2015). *Collective agency and cooperation in natural and artificial systems*. Cham: Springer International Publishing Switzerland.
Doi:10.1007/978-3-319-15515-9_1
- Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3, 1-21.
Doi:10.1177/2053951716679679
- Müller-Hengstenberg, C. D., & Kirn, S. (2016). *Rechtliche Risiken autonomer und vernetzter Systeme: eine Herausforderung*. Berlin/Boston: Walter de Gruyter GmbH.
- NHTSA (2013). Preliminary statement of policy concerning automated vehicles. US National Highway Traffic Safety Administration.
- Nickerson, R., Muntermann, J., Varshney, U., & Isaac, H. (2009). Taxonomy development in information systems: Developing a taxonomy of mobile applications. <https://halshs.archives-ouvertes.fr/halshs-00375103/document>. Accessed 3 August 2019
- Nof, S. Y. (2009). Automation: What it means to us around the world. In S. Nof (Ed.) *Springer Handbook of Automation* (pp. 13-52). Berlin/Heidelberg: Springer.

-
- Nunes, I., & Jannach, D. (2017). A systematic review and taxonomy of explanations in decision support and recommender systems. *User Modeling and User-Adapted Interaction*, 27, 393-444. Doi:10.1007/s11257-017-9195-0
- Onnasch, L., Maier, X., & Jürgensohn, T. (2016). *Mensch-Roboter-Interaktion: Eine Taxonomie für alle Anwendungsfälle*. Dortmund: Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (BauA).
- Pagallo, U. (2017). From automation to autonomous systems: A legal phenomenology with problems of accountability. *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, 17-23. Doi:10.24963/ijcai.2017/3
- Parasuraman, R., Sheridan, T. B., Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 30, 286-297. Doi:10.1109/3468.844354
- Proud, R. W., Hart, J. J., Mrozinski, R. B. (2003). *Methods for determining the level of autonomy to design into a human spaceflight vehicle: a function specific approach*. Houston, NASA Johnson Space Center Report.
- Rammert, W. (2009). Hybride Handlungsträgerschaft: Ein soziotechnisches Modell verteilten Handelns. In O. Herzog & T. Schildhauer (Eds.), *Intelligente Objekte* (pp. 23-33), Berlin/Heidelberg: Springer.
- Rammert, W., & Schulz-Schaeffer, I. (2002). Technik und Handeln: wenn soziales Handeln sich auf menschliches Verhalten und technische Artefakte verteilt. In W. Rammert & I. Schulz-Schaeffer (Eds.), *Können Maschinen handeln?: soziologische Beiträge zum Verhältnis von Mensch und Technik* (pp. 11-64). Frankfurt am Main: Campus Verlag.
- Rijsdijk, S. A., Hultink, E. J., & Diamantopoulos, A. (2007). Product intelligence: its conceptualization, measurement and impact on consumer satisfaction. *Journal of the Academy of Marketing Science*, 35, 340-356. Doi:10.1007/s11747-007-0040-6
- Riley, V. (1989). A general model of mixed-initiative human-machine systems. *Proceedings of the human factors society annual meeting*, 33, 124-128. Doi:10.1177/154193128903300227
- Russell, S. J., & Norvig, P. (2014). *Artificial intelligence: a modern approach*. Malaysia: Pearson Education Limited.

- Santosuosso, A., & Bottalico, B. (2017). Autonomous systems and the law: Why intelligence matters. In E. Hilgendorf & U. Seidel (Eds.), *Robotics, autonomies, and the law* (pp. 27-58). Baden-Baden: Nomos.
- Sartor, G., & Omicini, A. (2016). The autonomy of technological systems and responsibilities for their use. In N. Bhuta, S. Beck, R. Geiss, H. Y. Lui, & C. Kress (Eds.), *Autonomous weapon systems: law, ethics, policy* (pp. 39-74). New York/Cambridge: Cambridge University Press.
- Sheridan, T. B., & Verplank, W. L. (1978). Human and Computer Control of Undersea Teleoperators. Cambridge: Institute of Technology Cambridge. <http://www.dtic.mil/dtic/tr/fulltext/u2/a057655.pdf>. Accessed May 23 2019
- Shin, D., & Park, Y. J. (2019). Role of fairness, accountability, and transparency in algorithmic affordance. *Computers in Human Behavior*, 98, 277-284. Doi:10.1016/j.chb.2019.04.019
- Shneiderman, B. (2016). The dangers of faulty, biased, or malicious algorithms requires independent oversight. *Proceedings of the National Academy of Sciences of the United States of America*, 113, 13538-13540. Doi:10.1073/pnas.1618211113
- Simmler, M. (2019). Maschinenethik und strafrechtliche Verantwortlichkeit. In O. Bendel (Ed.), *Handbuch Maschinenethik* (pp. 1-18). Wiesbaden: Springer.
- Sommerville, I. (2007). *Software Engineering*. Essex: Pearson Education Limited.
- Thürmel, S. (2015). The participatory turn: A multidimensional gradual agency concept for human and non-human actors. In C. Misselhorn (Ed.), *Collective agency and cooperation in natural and artificial systems* (pp. 45-60). Cham: Springer.
- Vagia, M., Transeth, A. A., & Fjerdings, S. A. (2016). A literature review on the levels of automation during the years. What are the different taxonomies that have been proposed? *Applied Ergonomics*, 53, 190-202. Doi:10.1016/j.apergo.2015.09.013
- Verhagen, H. (2004). Autonomy and Reasoning for Natural and Artificial Agents. In M. Nickles, M. Rovatsos, & G. Weiss (Eds.), *Agents and computational autonomy. Lecture notes in computer science* (pp. 83-94). Berlin/Heidelberg: Springer.

-
- Wein, L. E. (1992). Responsibility of intelligent artifacts: Toward an automation jurisprudence. *Harvard Journal of Law & Technology*, 6, 103-154.
<https://heinonline.org/HOL/P?h=hein.journals/hjlt6&i=109>. Accessed 8 August 2019
- Weyer, J. (2006). *Die Kooperation menschlicher Akteure und nicht-menschlicher Agenten: Ansatzpunkte einer Soziologie hybrider Systeme* (Working Paper, 16-2006). Wirtschafts- und Sozialwissenschaftliche Fakultät Universität Dortmund. <http://nbn-resolving.de/urn:nbn:de:0168-ssoar-120992>.
- Weyer, J. & Reineke, S. (2005). *Creating order in hybrid systems: Reflections on the interaction of man and smart machines* (Working Paper, 7-2005). Technische Universität Dortmund. <http://nbn-resolving.de/urn:nbn:de:0168-ssoar-109749>.
- Zarsky, T. (2016). The trouble with algorithmic decisions: An analytic road map to examine efficiency and fairness in automated and opaque decision making. *Science Technology & Human Values*, 41, 118-132.
 Doi:10.1177%2F0162243915605575

Article II – A social cognition perspective on technical autonomy

Ruth Frischknecht

Abstract

The last several years have been characterized by an increasing autonomization of technology. Advanced technologies are now capable of learning and interacting. These capabilities expand the scope in which technology can be used. As a result, these increasingly autonomous machines and programs are completing an ever-growing number of tasks in collaboration with humans. This is why it is important to determine whether increasing levels of autonomy affect how technical systems are perceived by their human counterparts. An online survey experiment using vignettes to describe technical systems with varying levels of autonomy was employed to test the effects of technical autonomy on the dimensions of social cognition (i.e., warmth and competence) and mind perception (i.e., experience and agency). The findings show that increasing levels of technical autonomy impacted how the technical systems were socially perceived. More specifically, the findings suggest that the more autonomous a technical system is, the more competence and agency it will be ascribed. Conversely, increasing levels of technical autonomy had no effect on the ascription of warmth and experience. These results highlight that technical autonomy is relevant to and effectively shapes social judgements regarding technology.

Keywords: Human-computer interaction; technical autonomy; social cognition; mind perception; survey experiment

1 Introduction

Advanced technical systems are widespread in today's world taking over more and more tasks in our society. For example, advanced algorithms are used in personnel selection (van Esch, Black, & Ferolie, 2019) and the assessment of offenders (Dressel & Farid, 2018). Social robots care for the elderly and educate children (Broadbent, 2017). Their ability to complete such complex tasks changes the relationship between humans and machines. Machines are no longer mere tools (Rammert, 2011), but rather a type of companion (de Visser, Pak, & Shaw, 2018) or teammate (McNeese, Demir, Cooke, & Myers, 2018; Seeber, Bittner, et al., 2020). Their role has changed from one of decision supporter to actual decision maker (Nissen & Sengupta, 2006).

For technical systems to fulfill such complex tasks, machines must be able to respond to situations for which they have not been preprogrammed (Araujo, Helberger, Kruikemeier, & de Vreese, 2020). As a result, such systems are described as becoming more intelligent and autonomous (Corti & Gillespie, 2016; Seeber, Bittner, et al., 2020). Although there is no universal definition of technical autonomy per se, scholars seem to agree on some of its prerequisites. Adaptability and machine learning are considered core features of autonomous technology (de Visser et al., 2018; Jordan & Mitchell, 2015; Seeber, Waizenegger, et al., 2020). The capacity to interact with and enlarge their source data is another feature of technical autonomy (de Visser et al., 2018; Jordan & Mitchell, 2015; Seeber, Waizenegger, et al., 2020). Yet this does not mean that every technical system is equipped with the characteristics that shape technical autonomy. Rather, current developments include different technical systems that operate on various levels of technical autonomy. This means that technical autonomy is not to be understood as a binary category (i.e. a technical system is either autonomous or non-autonomous) but that it is a gradual property. According to this view, technical autonomy consists of different levels (de Visser et al., 2018; Endsley, 2017; Simmler & Frischknecht, 2021) or as comprised of different dimensions (Beer, Fisk, & Rogers, 2014). This suggests that there can be different

types or levels of technical autonomy. Endsley (2017), for example, introduced the notion of “semi-autonomous systems” to indicate that technology today does not reach full autonomy, nor is it strictly determined. Simmler and Frischknecht (2021) proposed a taxonomy for capturing technical autonomy that is comprised of five consecutive levels, providing an even more precise conception of how increasing technical autonomy can be mapped and described. It seems that the understanding of what technical autonomy is, is currently becoming clearer and more nuanced. However, the emphasis of the work cited above lies mainly on technology or put differently, on the machine part of the in human-machine collaboration.

The implementation of increasingly autonomous technical systems is associated with a wide array of social questions (Morris, Schlenoff, & Srinivasan, 2017; Rahwan et al., 2019). This calls for a more thorough understanding of how people respond to increasingly autonomous technologies, especially as the actions of autonomous technical agents become increasingly more intractable, opaque and difficult to comprehend (Rahwan et al., 2019). For example, it is important to explore the extent to which increasingly autonomous machines are seen as responsible for their actions (Awad et al., 2020; Banks, 2019) and whether decisions made by autonomous algorithms are seen as trustworthy or fair (Hoddinghaus, Sondern, & Hertel, 2021; Lee, 2018). These studies highlight that it not only matters what the machine is objectively capable of, but rather how people subjectively perceive technical system and their actions. This is further illustrated by Lee (2018), who found that the actions of an algorithm mattered as much as what people thought the algorithm was capable of doing when those people were asked to judge its trustworthiness and fairness. From this, it follows that people’s perceptions of technical systems play an important role in human-machine interactions. In fact, the significance of such perceptions has been emphasized in the literature on human-machine (Araujo et al., 2020; Lee, 2018; Sundar, 2020) and human-robot interaction. Despite the recognition that people’s perceptions of technology are relevant and important, little is known regarding if and how increasing levels of autonomy shape such perceptions.

Therefore, this research aims to fill this gap by systematically exploring *whether and how technical autonomy shapes social perceptions*.

2 The social perception of technology

Various studies have shown that humans treat, perceive, and interact with technologies as if they are social actors (for an overview see e.g. Reeves, 1996). For example, research has found that people apply social norms such as politeness when interacting with computers (Nass, Moon, & Carney, 1999) and rely on gender stereotypes (Nass, Moon, & Green, 1997). Initial work in this area focused on computers, technical systems unable to learn or increase their source data, and thus machines that operate at a rather low level of autonomy. These findings have been replicated with more sophisticated technical agents such as chatbots and robots, in the former offering evidence that people tend to interact with them as if they are human (Corti & Gillespie, 2016) and in the latter categorizing them according to gender stereotypes (Eyssel & Hegel, 2012). Thus, people react socially to technical agents, regardless of whether they are simple computers or advanced chatbots.

This tendency to perceive and treat non-human entities such as computers and robots as if they are human is called anthropomorphism (Epley, Waytz, & Cacioppo, 2007). Anthropomorphizing non-human entities is a highly adaptable heuristic for making sense of one's environment (Epley et al., 2007). Children and adults alike describe a wide range of non-human entities and phenomena (such as animals or the weather) in anthropomorphic ways (Zlotowski, Proudfoot, Yogeewaran, & Bartneck, 2015). Voice (Schroeder & Epley, 2016), movement patterns (Broadbent, 2017), and facial cues (Kim, Zhang, & Park, 2018) have all been found to reinforce the tendency to anthropomorphize technical agents. However, anthropomorphism only appears on an implicit level; when participants were asked explicitly if they would ascribe human-like characteristics to robots, they said they would not (Zlotowski et al., 2015). Thus, anthropomorphizing technical agents is

considered an unintentional process that happens automatically (Epley et al., 2007). From this, it follows that people tend to attribute human characteristics to technical systems and treat technical systems like social actors. For this reason, the dimensions of social cognition, as well as the dimensions of mind perception, are worthy of closer examination because both capture social perceptions in a very basic manner.

In the field of social cognition, it is well-recognized that people perceive other entities, usually other people or groups of people, in terms of warmth and competence; together, these perceptions comprise the universal dimensions of social cognition (Fiske, 2018; Fiske, Cuddy, & Glick, 2007). The warmth dimension describes the perceived intent of an entity and includes ascriptions of friendliness and helpfulness, whereas the competence dimension describes the perceived ability of an entity and includes ascriptions of intelligence and efficacy (Fiske et al., 2007). Findings in the field of social cognition research have already been applied to human-machine interactions. The Robotic Social Attributes Scale, an instrument for measuring people's judgment of robots, draws upon these two universal dimensions of social cognition (Carpinella, Wyman, Perez, & Stroessner, 2017). However, no general pattern of warmth and competence ascriptions to technical systems can be inferred from the literature. On the one hand, there is evidence that robots are ascribed more competence than warmth. On the other hand, Kim, Schmitt & Thalmann (2019) found that anthropomorphic features affect the judgment of warmth but not of competence. Ascriptions of warmth have also been found to predict trust in human-machine interactions. The more warmth a technical system is ascribed, the more trustworthy it appears (Kulms & Kopp, 2018). A study explicitly investigating the impact of the perceived level of technical autonomy on the social perception of intelligent personal assistants (such as Amazon's Alexa and Apple's Siri) suggested that increased autonomy increases ascriptions of both warmth and competence (Hu, Lu, Pan, Gong, & Yang, 2021). In sum, these findings suggest that people tend to ascribe warmth and competence to

technical systems and that increasing levels of technical autonomy may lead to more ascribed warmth and competence.

Another approach to understanding social perception centers on a more specific domain: the mind. Researchers have investigated if and under what conditions an entity is perceived as having a mind. According to this paradigm, mind perception is comprised of two dimensions: experience and agency (Gray, Gray, & Wegner, 2007; Waytz, Gray, Epley, & Wegner, 2010). While experience describes whether an entity is ascribed the capacity to sense and feel, agency indicates if an entity is ascribed the ability to plan and act (Waytz, Gray, et al., 2010). According to Gray et al. (2007) mind perception is closely interlinked with an entity's moral status: Someone (or something) perceived as having substantial agency is considered a moral agent, and as such bears moral responsibility. Inversely, entities perceived as being capable of experience are regarded as moral patients, and as such possess moral rights (Waytz, Gray, et al., 2010).

In their study of mere ascription of mind perception dimensions Gray et al. (2007) compared different entities with regards to how much and what type of mind they were ascribed. This comparison showed that human beings are attributed a considerable amount of agency and experience. An animal such as a frog is attributed little agency but much experience. In that study, participants were also asked to rate a social robot, which was ascribed a medium amount of agency and almost no experience. Thus, people tend to ascribe mind perception dimensions to technical systems. The ascription of a mind is closely interlinked with contextual factors such as the function a technical system fulfills. If a technical system primarily performs social tasks and acts as a caregiver, it is perceived as having a stronger capacity to experience (as opposed to a condition where the same system primarily performs economic tasks and acts as a salesperson) (Wang & Krumhuber, 2018). In an experiment with a self-driving car, participants ascribed more agency and experience to a vehicle that was given a name, voice, and gender, as compared to a self-driving car without these cues (Waytz, Heafner, & Epley, 2014). Additionally, (un)predictability has been

found to be a determinant in ascribing a mind; the more unpredictable the entity acts, the more mindful it appears (Waytz, Gray, et al., 2010). Applied to the concept of technical autonomy, which renders actions of technical systems opaque, this suggests that an increasing level of autonomy might make the ascription of a mind more likely.

2.1 The present study

Overall, the findings reported above show that people readily ascribe to technical systems concepts and variables that are usually employed in human-human perception, such as warmth, competence, agency, and experience. In studies concerned with mind perception, these ascriptions were found to follow a persistent pattern; technical systems were ascribed more agency (and thus the capacity to act) than experience (and thus the capacity to sense) (Gray et al., 2007; Stafford, MacDonald, Jayawardena, Wegner, & Broadbent, 2014). However, none of the studies cited above systematically tested how increasing levels of technical autonomy shape such perception. The pattern of ascription related to the universal dimensions of social cognition (i.e., warmth and competence) is less clear. Some studies indicated that technical systems were ascribed competence but not warmth (Pan, Croft, & Niemeyer, 2018), whereas other findings suggested that under increasing levels of technical autonomy perceived warmth and competence increased (Hu et al., 2021). In other words, the current state of knowledge on the influence of increasing autonomy on the social perception of technology is both ambiguous and fragmentary. Likewise, the notion that technical autonomy is gradual and different from automation is relatively new (see, e.g., Simmler & Frischknecht, 2021). Rather, over the past several decades, technical systems have been treated as constants (Sundar, 2020), and generally, the perception has been that a simple computer tends to elicit more or less the same social response as a sophisticated virtual assistant (see e.g., Nass & Moon, 2000; Reeves, 1996).

Thus, the goal of the present study is to further explore how *increasing levels of technical autonomy* generally affect such ascriptions. To do so, this research systematically explores *if* and *how* increasing levels of technical autonomy affect the social perception of a technical system. We hypothesize that the level of technical autonomy shapes how the technical system is socially perceived, and thus affects how much warmth, competence, experience, and agency it is ascribed. We further expect that the more autonomous the technical system is, the more warmth, competence, experience, and agency it will be ascribed.

To experimentally test the effect of increasing levels of autonomy on social perception, vignettes were developed for the different groups that illustrated varying levels of autonomy. To do so, the approach of Simmler and Frischknecht (2021) was employed because it provided a straightforward description of five different, consecutive levels of technical autonomy. According to their research, technical autonomy is mainly shaped by four dimensions: indetermination, non-transparency, adaptability, and openness. A technical system that is completely determined and transparent, unable to adapt to its environment, and not open (i.e., closed) is said to have no autonomy. Conversely, an indetermined, non-transparent, adaptable, and open system is considered to reach the highest level of technical autonomy. In between lie three other levels of autonomy, or what they called non-transparent, indetermined, and adaptable systems. A more detailed description of the five possible levels of technical autonomy and what technical features delineates them can be found in Table 7. It is important to note that the taxonomy represents autonomy in terms of today's possibilities, and further developments might allow for even higher levels of technical autonomy.

Table 7: Five levels of technical autonomy and main characteristics (Simmmler & Frischknecht, 2021)* and their correspondence in the vignette.

Level of Technical Autonomy	Features of Technical Autonomy	Correspondence with the Vignette
Deterministic system (no technical autonomy, level 0)	determined, transparent, non-adaptable, closed	completely predetermined, the same input always leads to the same output, easily traceable, neither learns nor adapts, the data are specified, the system is closed
Non-transparent system (level 1)	determined, <i>non-transparent</i> , non-adaptable, closed	completely predetermined, the same input always leads to the same output, <i>not easily traceable</i> , neither learns nor adapts, the data are specified, the system is closed
Indetermined system (level 2)	<i>indetermined</i> , non-transparent, non-adaptable, closed	<i>not completely predetermined</i> , <i>the same input does not always lead to the same output</i> , not easily traceable, neither learns nor adapts, the data are specified, the system is closed
Adaptable system (level 3)	indetermined, non-transparent, <i>adaptable</i> , closed	not completely predetermined, the same input does not always lead to the same output, not easily traceable, <i>able to learn and adapt</i> , the data are specified, the system is closed
Open system (level 4)	indetermined, non-transparent, adaptable, <i>open</i>	not completely predetermined, the same input does not always lead to the same output, not easily traceable, able to learn and adapt, <i>the data are not specified</i> , <i>the system is open</i>

*Property changes leading to higher levels of technical autonomy are marked in italics.

3 Method

This study employed a one-factorial between-subject design. Technical autonomy was the single factor and consisted of five different levels. In order to anchor the impressions of the technical system, a condition describing a human being was added. This resulted in six different experimental conditions.

3.1 Participants

All participants ($N = 300$) were recruited via Amazon's Mechanical. M-turk samples have been proven to be valid and of similar quality as data collected by more conventional means (Buhrmester, Kwang, & Gosling, 2011; Casler, Bickel, & Hackett, 2013; Paolacci, Chandler, & Ipeirotis, 2010). To ensure high data quality, participants were selected based on three criteria. They had to be US residents, have a Human Intelligence Task (HIT) approval rate greater than 98%, and completed more than 500 HITs. Participants were paid \$1.20 for taking part in the study.

Participants were randomly assigned to one of the six conditions (consisting of the five levels of autonomy and one human control condition, see Table 7), resulting in 50 participants per condition. 12 participants had to be excluded because they failed at least one attention check. This resulted in a final sample of 288 participants (120 females, 167 males, and 1 person reporting "other" for gender). The mean age was 37.09 years ($SD = 11.41$).

3.2 Materials

3.2.1. Manipulation. Depending on the condition the participant was assigned, they received a description either of a technical system or human being. The descriptions were constructed based on the conception of technical autonomy provided by Simmler and Frischknecht (2021). Four technical characteristics were considered relevant for technical autonomy (i.e., transparency, determinism, adaptability, and openness; these were varied among the

conditions). The structure of each description was held constant. Only the system's characteristics were modified, and the rest of the description was held as constant as possible. The key characteristics were marked in bold font (as indicated in the sample vignette below). Participants received the following vignette, with the description matching the level of technical autonomy according to the assigned condition.

The actions of the technical system are **[completely/not completely] predetermined**, which means that the same input **[always/not necessarily]** leads to the same output. The steps to get from input to output **[are/are not] easily traceable** and transparent to an outsider/observer. The technical system **[can neither/is able to] learn (n)or adapt** to the environmental conditions in which it is embedded. The data upon which the system is based **[are/are not] specified in advance**. Therefore, the system is **[closed/open]**.

Table 7 depicts the five conditions in greater detail and illustrates how the manipulation fit the respective levels of autonomy proposed by Simmler and Frischknecht (2021). It is important to note that these descriptions were deliberately kept generic, text-based, and abstract because the goal of the experiment was to *generally* explore social perceptions as a function of technical autonomy. By providing a bare and abstract description of the technical system with regards to its capabilities, it is possible to isolate the effects of technical autonomy. In order to do so, appearance (Kim et al., 2018; Schroeder & Epley, 2016) and task domain (Wang & Krumhuber, 2018) had to be experimentally controlled by holding them constant, as both could have impacted participants' social perceptions. Put differently, by describing a technical system fulfilling a task such as caring for the elderly or including images or videos of a particular technical system, the effect of the technical autonomy could potentially have been confounded.

For exploratory reasons, a human (control) condition was also included in the experiment. This condition served to classify and contextualize the results.

Participants in this condition received the following instruction: “Think of human beings in general. Try to imagine their capabilities and characteristics.”

3.2.2. Dependent variables. All responses concerning the dependent variables were made on a 5-point Likert scale ranging from 1 (not at all) to 5 (very much). All items were presented in a randomized order to counterbalance any possible presentations effects. The survey measured four dependent variables: warmth and competence (together comprising the universal dimensions of social cognition) and experience and agency (together comprising the two dimensions of mind perception). The universal dimensions of social cognition, warmth and competence, were operationalized with the items suggested by Fiske (2018). The competence scale consisted of the following items: *competent, intelligent, skilled, efficient, assertive, and confident* (Cronbach’s $\alpha = 0.84$). The warmth scale was composed of the following items: *warm, trustworthy, friendly, honest, likable, and sincere* (Cronbach’s $\alpha = 0.89$). The mind perception dimensions were operationalized according to Wang and Krumhuber (2018). The agency scale consisted of the following items: *exercise self-control, think analytically, be rational, remember, and plan* (Cronbach’s $\alpha = 0.87$). The experience dimension was measured with the following items: *experience emotions, have feelings, be emotional, desire, and understand others* (Cronbach’s $\alpha = 0.98$).

3.2.3. Control variables. Possible covariates were also included in the survey experiment. The personal tendency to anthropomorphize served as the first control variable and was measured with the questionnaire developed by Waytz, Cacioppo, and Epley (2010) (Cronbach’s $\alpha = 0.79$). The instrument contained items such as “To what extent does a cheetah experience emotions?” Controlling for the personal tendency to anthropomorphize was important, it was possible that the stronger this tendency was, the more the participant would ascribe dimensions of social cognition and mind perception variables. Technical affinity, the second control variable, was operationalized with items proposed by Franke, Wessel, and Attig (2018) (Cronbach’s $\alpha = 0.82$). This included items such as “I try to understand precisely how a technical system works.” Including technical affinity allowed the possible confounding effects of participants’

interest in and experience with technical gadgets to be controlled, as knowledge and online self-efficacy have been found to be relevant determinants in technology perception (Araujo et al., 2020).

3.2.4. Manipulation and attention checks. In the manipulation check, participants were asked how autonomous, independent, self-sufficient, and self-reliant the system was according to their point of view. These items were employed in the self-developed perceived autonomy scale (Cronbach's $\alpha = 0.9$). A total of three attentions checks were built in between the dependent and control variables to identify inattentive participants. These three items took the form of directed queries (sample item: "In this row, click <undecided>"), as suggested by Abbey and Meloy (2017). The answer was considered correct if the participant clicked what was indicated in the query and incorrect if they clicked something else.

3.3 Procedure

The survey experiment was completed online and programmed on Unipark. After a general introduction, participants were provided with the following instructions on how to complete the survey: *Below, you will receive a description of a technical system and how it functions. Please read it carefully and make sure you understand the system's features. Of course, you may read the description more than once. When reading, please imagine this system in as much detail as possible.* Then the participants received a description of a technical system, depending on the condition to which they were randomly assigned. The most important features of the technical system were emphasized in bold font. In the human control condition, participants were instructed to simply rate their impressions of human beings in general. Then, participants received the following statement: *You have just read a description of a technical system. Now, you must evaluate it. Note that there is no correct answer. Simply rate the items according to your impressions.* The participants then rated their impressions regarding the universal dimensions of social cognition: competence and warmth. Next, they were instructed to read the same description of the technical system (or the human being) again and rate items related to the dimensions of mind perception:

agency and experience. Then, participants were instructed to complete the manipulation check items and asked to briefly describe what kind(s) of technical system(s) they had in mind when completing the survey. The two control variables of technical affinity and personal tendency to anthropomorphize were then measured. Lastly, participants provided demographic information and were given the opportunity to provide general feedback. Finally, the participants were thanked and received a payment code.

4 Results

4.1 Preliminary Analyses

All scales were found to be sufficiently reliable; therefore, all items were used for further analysis. In order to test if the manipulation worked (i.e., if the vignettes did in fact manipulate technical autonomy), a one-way analysis of variance (ANOVA) was conducted of the five levels of technical autonomy (ranging from deterministic to open system) as factors, and the amount of perceived autonomy as a dependent variable. Because the assumption of equal variances was violated, a robust one-way ANOVA using Welch's procedure revealed a statistically significant main effect ($F(4, 116.4) = 3.21, p = .015, \eta_p^2 = .047$). This suggests that the vignettes in fact manipulated technical autonomy.

4.2 Effects of technical autonomy

To explore the effects of technical autonomy on the general social perception of a technical system, a multivariate analysis of variance (MANOVA) was conducted of the levels of technical autonomy as independent variables and ascriptions of warmth, competence, agency, and experience as dependent variables. The group means and standard deviations can be found in Table 8. In order to check if the MANOVA was eligible, a series of Pearson's correlations were performed among all the dependent variables to see if they correlated in a moderate fashion.

Table 8: Descriptive statistics of the dependent variables.

	Warmth		Competence		Experience		Agency	
	M	SD	M	SD	M	SD	M	SD
Deterministic system	2.51	0.92	3.22	0.96	1.28	0.59	2.47	1.04
Non-transparent system	2.42	0.81	3.15	0.86	1.38	0.85	2.57	0.91
Indetermined system	2.24	0.85	2.82	0.85	1.37	0.73	2.39	1.21
Adaptable system	2.29	0.81	3.16	1.05	1.34	0.70	2.96	1.06
Open system	2.51	0.88	3.41	0.90	1.40	0.63	3.20	1.01
Human	4.28	0.74	4.45	1.02	4.89	0.26	4.49	0.68

The correlation pattern depicted in Table 9 suggests that a not to be statistically significant ($M = 53.37$, $p = .11$), indicating that the covariance matrices were roughly equal among the groups. As hypothesized, the MANOVA revealed a statistically significant multivariate effect of technical autonomy, Wilk's $\lambda = .86$, $F(16, 703) = 2.21$, $p = .004$, $\eta_p^2 = .037$. This statistically significant effect persisted when including the personal tendency to anthropomorphize and technical affinity as covariates, Wilk's $\lambda = .86$, $F(16, 697) = 2.25$, $p = .003$, $\eta_p^2 = .039$.

Table 9: Correlations among dependent variables.

	M	SD	1	2	3	4
1. Warmth	2.39	0.85	1			
2. Competence	3.15	0.94	.68	1		
3. Experience	1.35	0.70	.53	.33	1	
4. Agency	2.72	1.09	.49	.63	.39	1

*Note that the human condition was excluded from this analysis as it was not part of the MANOVA.

To follow up this analysis, the univariate effects of technical autonomy on every dependent variable were explored separately. The covariates remained in the model in order to further control for their effects. Before doing so, a Levene's test was run, which suggested that the assumption of homogeneity of the variances was met (smallest $p = .38$). On the one hand, these univariate analyses revealed a statistically significant main effect of technical autonomy on competence ($F(4, 231) = 2.22, p = 0.042, \eta_p^2 = .042$). Follow-up tests revealed that participants perceived the open system as significantly more competent ($M = 3.39, SE = .128, p = 0.023$) than the indeterminated system ($M = 2.80, SE = .126$). Regarding the deterministic ($M = 3.19, SE = 0.126$), non-transparent ($M = 3.12, SE = .125$), and adaptable systems ($M = 3.22, SE = .126$), no differences were found. This is in line with the hypothesis, according to which technical autonomy has an effect on the ascription of competence. However, the assumption that the higher the level of autonomy is, the more competence is ascribed received only partial support.

Another statistically significant univariate main effect was found on agency ($F(4, 231) = 6.09, p < 0.001, \eta_p^2 = .095$). Follow-up tests using the Bonferroni adjustment showed that the open system ($M = 3.20, SE = .148$) was ascribed statistically significant more agency than were the deterministic system ($M = 2.44, SE = .146, p = .004$), the non-transparent system ($M = 2.54, SE = .14, p = .017$), and the indeterminated system ($M = 2.39, SE = .149, p = .002$). Additionally, the adaptable system ($M = 3.02, SE = .146$) was ascribed statistically significant more agency than the indeterminated system ($M = 2.39, SE = .149, p = .031$). Thus, this is in line with the hypothesis that technical autonomy affects the ascription of

agency and supports the notion that the more autonomous a technical system is the more agency it is ascribed in tendency.

On the other hand, the univariate analyses showed no statistically significant main effect of technical autonomy on warmth ($F(4, 231) = 0.7, p = 0.6, \eta_p^2 = .012$) and on experience ($F(4, 231) = 0.36, p = .84, \eta_p^2 = .006$), indicating that the group means did not differ in a statistically meaningful manner. Follow-up tests using the Bonferroni correction revealed no differences among any of the experimental conditions. Thus, the findings regarding the ascription of warmth and experience both were not in line with the hypotheses. For better clarity, Figure 2 depicts the mean ascription of warmth, competence, experience, and agency across the five experimental conditions.

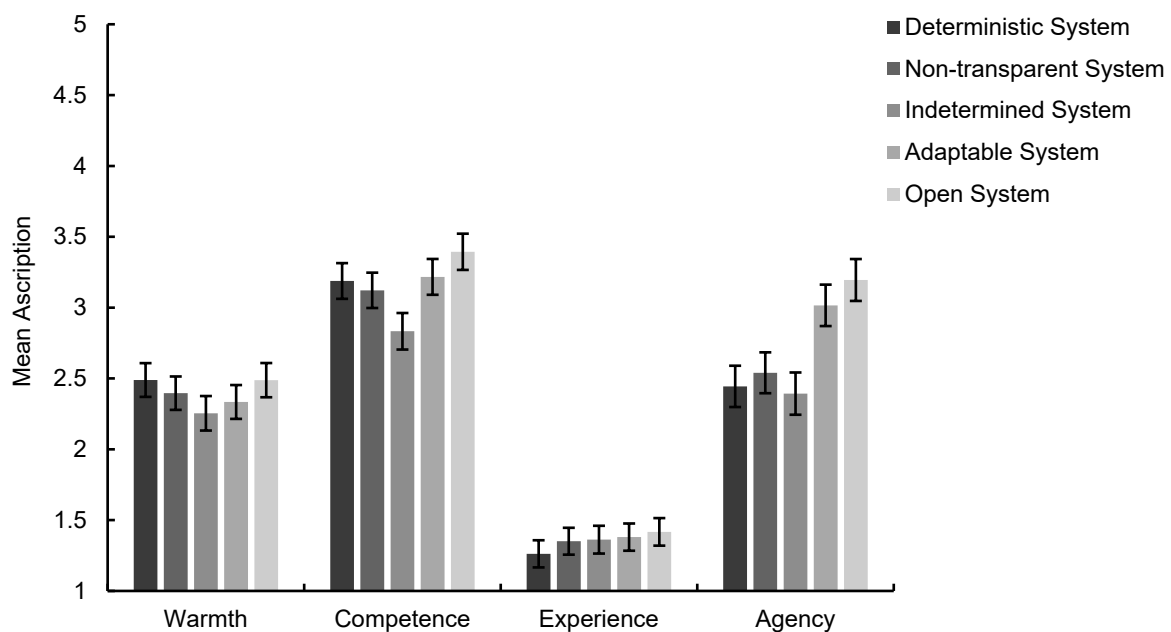


Figure 2: Mean ascriptions of social cognition and mind perception dimensions. Mean scores adjusted for covariates, anthropomorphism, and technical affinity by condition. Error bars represent standard errors.

4.3 Contrasting technical systems with humans

As explained above, for exploratory reasons, a human control group was included in the experimental design. However, it became evident that the human control condition differed drastically (and thus statistically significantly) from the conditions capturing technical autonomy; this was true for every dependent variable and the manipulation check. Thus, the condition was excluded from the main analyses, as the experiment was designed to explore the social perception of autonomous technologies and not to compare autonomous technical systems with humans. Because the assumption for MANOVA was violated, a Kruskal Wallis Test was conducted. This test revealed, a significant effect of technical autonomy on all dependent variables (i.e. warmth, competence, agency and experience), largest $p < .001$.

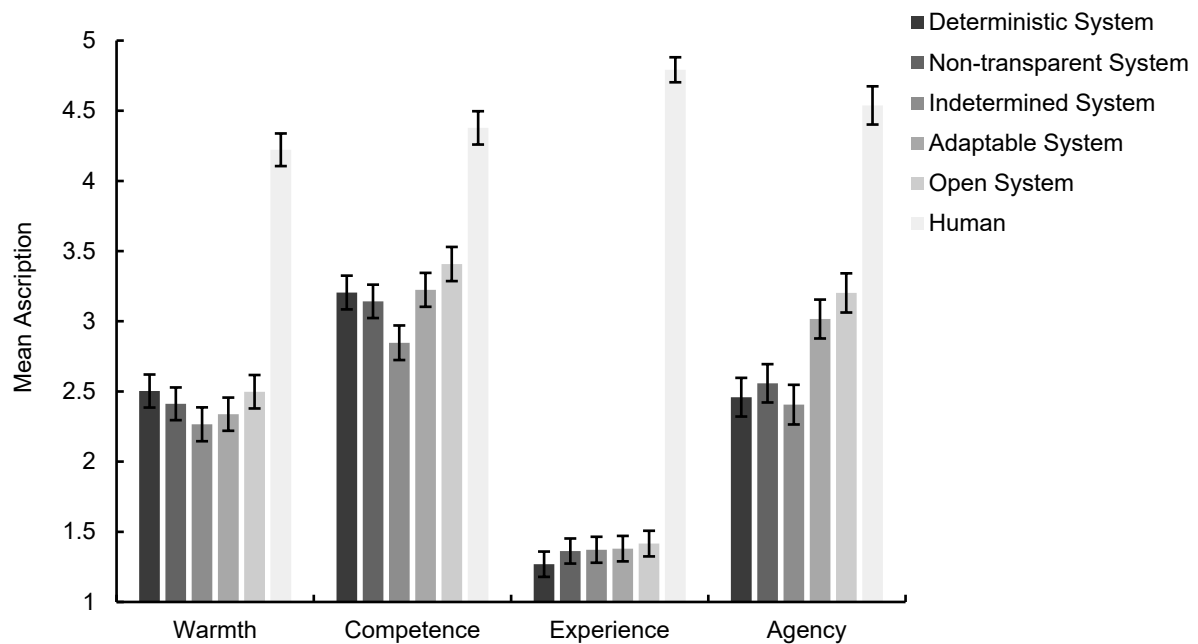


Figure 3: Mean ascriptions of social cognition and mind perception including the human control condition. Mean scores adjusted for covariates, anthropomorphism, and technical affinity by condition, including the human control condition. Error bars represent standard errors.

Additionally, follow-up tests using the Dunn-Bonferroni adjustment revealed that the human control condition differed statistically significantly from every level of technical autonomy in all dependent variables (all differences below $p < .001$). Figure 3 further highlights the difference between the conditions.

5 Discussion

The experiment described above was designed to explore if increasing levels of technical autonomy mattered for social perception, and if so, how increasing levels of technical autonomy might shape these social perceptions. Ascription patterns were systematically tested by operationalizing technical autonomy with theoretically grounded vignettes and thus describing technical systems operating on different levels of autonomy. The results suggest that increasing levels of technical autonomy did indeed influence the social perception of the technical systems. In particular, increasing levels of technical autonomy changed the ascription of competence and agency, while ascriptions of warmth and experience did not change under conditions of increasing technical autonomy. Furthermore, the results presented above were neither due to personal tendency for anthropomorphizing nor to technical affinity, as they were experimentally controlled due to their potentially confounding effects. It is also important to note that technology, even at a high level of autonomy, was attributed significantly less warmth, competence, experience, and agency than human beings.

In a broad sense, the results suggest *that* increasing levels of technical autonomy affect the social perception of technology in general. The evidence presented above suggests that technical autonomy shapes how technical systems are socially perceived, and therefore is an important point to consider when investigating human-machine interactions. Past research has argued that factors such as embodiment (Hoffmann, Bock, & Pütten, 2018), voice (Waytz et al., 2014), gender (Eyssel & Hegel, 2012), and the social role (Wang & Krumhuber, 2018) a technical system is assigned all matter with regards to how it is socially

perceived. Technical autonomy complements these findings by emphasizing that it may also be important to consider the level of autonomy at which the technical system operates.

In a narrower sense, the results also provide evidence of *how* the perception of technology is altered under conditions of increasing technical autonomy. While technical autonomy affected the ascription of competence and agency, it had no effect on ascriptions of warmth and experience. This pattern is in agreement with the results of past research on how technical agents are generally perceived: as more competent than warm as well as having more agency than experience (see for example Gray et al., 2007; Pan et al., 2018; Stafford et al., 2014).

Besides, the data contain evidence that ascriptions of competence and agency change under conditions of increasing levels of technical autonomy. The open system and the adaptable system, both operating on relatively high levels of autonomy, were ascribed more agency than technical systems operating at lower levels of autonomy. This difference thus became evident at the level where technical systems are described as capable of learning and adapting (see Simmler & Frischknecht, 2021). The significant increase of agency ascriptions thus might be explained by the adaptability of technical systems. Further investigating how machine learning algorithms shape social perception thus could be an interesting area for further research. Regarding the ascription of competence, a surprising pattern was observed. Although the open system was regarded as the most competent, the indeterministic systems considered the least competent. According to the taxonomy of Simmler and Frischknecht (2021) this means that it was not a system operating at a low level of autonomy but one at an intermediate level of autonomy that was ascribed the least competence. The data allow no direct conclusions about why this was the case. A possible explanation for this may be that the indeterministic system contains a random factor (Simmler & Frischknecht, 2021). This could have given the impression of an arbitrary or random system and thus might have had a negative impact on the ascription of competence.

Conversely, technical had no effect on the ascription of experience. Experience is closely related to having and expressing emotions (Gray et al., 2007) and thus may simply have been too far removed from what a technical system, regardless of its level of autonomy, can realistically be ascribed. Likewise, the findings indicate no effect of increasing levels of autonomy on the ascription of warmth, which is in contrast to the results of Hu et al. (2021). According to that study, the higher the perceived autonomy of an intelligent service assistant was, the more warmth participants ascribed to it. One potential explanation of why the present study failed to replicate these findings might lie in the operationalization of technical autonomy. While this study presented generic vignettes referring to “a technical system,” Hu et al. (2021) investigated the effects of intelligent virtual assistants, and thus technical agents with anthropomorphic features such as names and voices that fulfilled social tasks and helped or assisted human beings. Past research has shown that characteristics such as a human-like appearance increases perceptions of warmth (Kim et al., 2019), and that occupying a social role determines the ascription of experience (Wang & Krumhuber, 2018). Thus, it may be that technical autonomy per se impacts the ascription of competence and agency, while bodily or contextual features such as a friendly face or social role impact the ascription of warmth and experience. Because this study does not provide a potential explanation for this, future research should explore how technical autonomy interacts with other factors. A deeper understanding of how the perception of warmth and technical autonomy relate is crucial because the perception of warmth seems to determine trust (Kulms & Kopp, 2018), and trust is highly relevant in the context of human-machine interactions (de Visser et al., 2018).

As this study’s aim was to explore the impact of technical autonomy in only a general sense, no direct behavioral consequences of these ascriptions were measured. However, by matching the findings with results from past research, two possible implications become clear. First, competence and agency describe behavioral and intellectual abilities, and the results suggest that people ascribe

more intellectual ability to technical systems described as operating with higher levels of technical autonomy. Because of this, people may be more likely to act on suggestions and recommendations from systems operating at higher levels of autonomy. Yet, this remains only a proposition. In their series of experiments, Logg, Minson, and Moore (2019) showed that people in fact are willing to rely on the advice of algorithms. Future research should also investigate how increasing levels of technical autonomy affect this willingness, what actual impacts (if any) increasing levels of technical autonomy have on decisions, and if people do, in fact, tend to rely more heavily on outputs produced by systems with greater levels of autonomy. Secondly, in the present research, as the autonomy level increased, more agency was attributed to the technical system. Entities described as being capable of agency are considered moral agents, which are usually regarded as being responsible for their actions (Waytz, Gray, et al., 2010). So far, it has been shown that machines are blamed less for errors than are humans (Awad et al., 2020). However, it may be that judgments regarding responsibility change under conditions of increasing technical autonomy. Whether or not advanced technology is regarded as responsible for its actions could for example have implications for future legislation.

The experiment and findings presented here have certain limitations. Two of those limitations are related to the application of generic vignettes that abstractly described technical autonomy and did not provide any further information. First, this may have been difficult to understand. However, the manipulation check showed that perceived autonomy differed across the conditions and that the relationship was linear (i.e., the higher the level of autonomy described in the vignette, the more autonomous the technical system was perceived by the participants). This suggests that the manipulation of the level of autonomy did work as intended and the vignettes were sufficiently comprehensible. Second, the generic vignettes might limit the applicability of the results in more specific contexts. The perception of technology and human reactions to it might vary by contexts (Rahwan et al., 2019). For example, Wang and Krumhuber (2018) demonstrated that a robot was ascribed more experience

when it fulfilled a social role as opposed to when that same robot fulfilled an economic role. Given this context dependence of human-machine interactions, it might be that technical autonomy is perceived differently when it is implemented in a hospital rather than an investment bank. Thus, testing the effects of technical autonomy on social ascription patterns across different social contexts would be an interesting area of further research. Another concern is related to the external validity of the results. The data were gathered via an online experiment, and thus did not allow for predictions regarding the influence of increasing levels of technical autonomy in the real world. In the real world, human-machine interactions take place in complex settings where attention is never directed to a single feature of the technology. Thus, the experiment likely overestimated the effects of technical autonomy in the real world. Lastly, due to this experiment's explorative nature, possible moderators and mediators were not considered in the design. Likewise, it was beyond this paper's scope to test associations between warmth, competence, experience and agency and how the levels of autonomy shape combinations of these perceptions.

6 Conclusion

Self-driving cars, predictive policing software, and care robots all require advanced technical capabilities such as the ability to adapt to their environment and enlarge their initial source data (Rahwan et al., 2019). As a result, these technical systems have become more autonomous. This study explored the influence of increasing levels of autonomy on how machines are socially perceived, since people tend to treat technology as a social actor (Nass & Moon, 2000). For many years, the prevailing view was that people responded socially to technology and that functionality played little role in the response (i.e., a simple computer would elicit roughly the same social response as an advanced, adaptable recommender system) (Reeves & Nass, 1996). Technical functioning such as technical autonomy was considered a constant (Sundar, 2020) and thus regarded as negligible. Counter to this, the present work suggests that the functioning of a technology does, in fact, matter for social perception because under conditions of increasing technical autonomy, social perceptions differ and change. Technical autonomy is a relevant variable that matters in human-machine interactions.

7 References

- Abbey, J. D., & Meloy, M. G. (2017). Attention by design: Using attention checks to detect inattentive respondents and improve data quality. *Journal of Operations Management*, 53-56(1), 63-70. doi: 10.1016/j.jom.2017.06.00
- Araujo, T., Helberger, N., Kruikemeier, S., & de Vreese, C. H. (2020). In AI we trust? Perceptions about automated decision-making by artificial intelligence. *Ai & Society*, 35(3), 611-623. doi: 10.1007/s00146-019-00931-w
- Awad, E., Levine, S., Kleiman-Weiner, M., Dsouza, S., Tenenbaum, J. B., Shariff, A., . . . Rahwan, I. (2020). Drivers are blamed more than their automated cars when both make mistakes. *Nature Human Behaviour*, 4(2), 134-143. doi: 10.1038/s41562-019-0762-8
- Banks, J. (2019). A perceived moral agency scale: Development and validation of a metric for humans and social machines. *Computers in Human Behavior*, 90, 363-371. doi: 10.1016/j.chb.2018.08.028
- Beer, J. M., Fisk, A. D., & Rogers, W. A. (2014). Toward a framework for levels of robot autonomy in human-robot interaction. *Journal of Human-Robot Interaction*, 3(2), 18.
- Broadbent, E. (2017). Interactions With Robots: The Truths We Reveal About Ourselves. In S. T. Fiske (Ed.): *Annual Review of Psychology* (Vol. 68, pp. 627-652)
- Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? *Perspectives on Psychological Science*, 6(1), 3-5. doi: 10.1177/1745691610393980
- Carpinella, C. M., Wyman, A. B., Perez, M. A., & Stroessner, S. J. (2017). *The Robotic Social Attributes Scale (RoSAS): Development and Validation*. Paper presented at the Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction, Vienna, Austria.
- Casler, K., Bickel, L., & Hackett, E. (2013). Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing. *Computers in Human Behavior*, 29(6), 2156-2160. doi: 10.1016/j.chb.2013.05.009
- Corti, K., & Gillespie, A. (2016). Co-constructing intersubjectivity with artificial conversational agents: People are more likely to initiate repairs of misunderstandings with agents represented as human. *Computers in Human Behavior*, 58, 431-442. doi: <https://doi.org/10.1016/j.chb.2015.12.039>
- de Visser, E.J., Pak, R., & Shaw, T.H. (2018). From "automation" to "autonomy": the importance of trust repair in human-machine interaction. *Ergonomics*, 61(10), 1409-1427.

-
- Dressel, J., & Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism. *Science Advances*, 4(1). doi: 10.1126/sciadv.aao5580
- Endsley, M. R. (2017). From Here to Autonomy: Lessons Learned From Human-Automation Research. *Human Factors*, 59(1), 5-27. doi: 10.1177/0018720816681350
- Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review*, 114(4), 864-886. doi: 10.1037/0033-295x.114.4.864
- Eyssel, F., & Hegel, F. (2012). (S)he's Got the Look: Gender Stereotyping of Robots. *Journal of Applied Social Psychology*, 42(9), 2213-2230. doi: 10.1111/j.1559-1816.2012.00937.x
- Fiske, S. T. (2018). Stereotype Content: Warmth and Competence Endure. *Current Directions in Psychological Science*, 27(2), 67-73. doi: 10.1177/0963721417738825
- 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. doi: 10.1016/j.tics.2006.11.005
- Franke, T., Wessel, D., & Attig, C. (2018). A Personal Resource for Technology Interaction: Development and Validation of the Affinity for Technology Interaction (ATI) Scale. *International Journal of Human-Computer Interaction*, 1-12. doi: 10.1080/10447318.2018.1456150
- Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science*, 315(5812), 619-619. doi: 10.1126/science.1134475
- Hoffmann, L., Bock, N., & Pütten, A. M. R. v. d. (2018). *The Peculiarities of Robot Embodiment (EmCorp-Scale): Development, Validation and Initial Test of the Embodiment and Corporeality of Artificial Agents Scale*. Paper presented at the Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction, Chicago, IL, USA. <https://doi.org/10.1145/3171221.3171242>
- Hu, Q., Lu, Y. B., Pan, Z., Gong, Y. M., & Yang, Z. L. (2021). Can AI artifacts influence human cognition? The effects of artificial autonomy in intelligent personal assistants. *International Journal of Information Management*, 56, 15. doi: 10.1016/j.ijinfomgt.2020.102250
- Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255-260. doi: 10.1126/science.aaa8415
- Kim, S., Zhang, K., & Park, D. (2018). Don't Want to Look Dumb? The Role of Theories of Intelligence and Humanlike Features in Online Help Seeking. *Psychological Science*, 29(2), 171-180. doi: 10.1177/0956797617730595
- Kim, S. Y., Schmitt, B. H., & Thalmann, N. M. (2019). Eliza in the uncanny valley: anthropomorphizing consumer robots increases their perceived warmth

- but decreases liking. *Marketing Letters*, 30(1), 1-12. doi: 10.1007/s11002-019-09485-9
- Kulms, P., & Kopp, S. (2018). A Social Cognition Perspective on Human-Computer Trust: The Effect of Perceived Warmth and Competence on Trust in Decision-Making With Computers. *Frontiers in Digital Humanities*, 5(14). doi: 10.3389/fdigh.2018.00014
- Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*, 5(1), 2053951718756684. doi: 10.1177/2053951718756684
- 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. doi: 10.1016/j.obhdp.2018.12.005
- McNeese, N. J., Demir, M., Cooke, N. J., & Myers, C. (2018). Teaming With a Synthetic Teammate: Insights into Human-Autonomy Teaming. *Human Factors*, 60(2), 262-273. doi: 10.1177/0018720817743223
- Morris, K. C., Schlenoff, C., & Srinivasan, V. (2017). A Remarkable Resurgence of Artificial Intelligence and Its Impact on Automation and Autonomy. *Ieee Transactions on Automation Science and Engineering*, 14(2), 407-409. doi: 10.1109/tase.2016.2640778
- Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, 56(1), 81-103. doi: 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), 864-876. doi: 10.1111/j.1559-1816.1997.tb00275.x
- Nass, C., Moon, Y. M., & Carney, P. (1999). Are people polite to computers? Responses to computer-based interviewing systems. *Journal of Applied Social Psychology*, 29(5), 1093-1110. doi: 10.1111/j.1559-1816.1999.tb00142.x
- Nissen, M. E., & Sengupta, K. (2006). Incorporating Software Agents into Supply Chains: Experimental Investigation with a Procurement Task. *MIS Quarterly*, 30(1), 145-166. doi: 10.2307/25148721
- Pan, M., Croft, E. A., & Niemeyer, G. (2018). Evaluating Social Perception of Human-to-Robot Handovers Using the Robot Social Attributes Scale (RoSAS) *Hri '18: Proceedings of the 2018 Acm/Ieee International Conference on Human-Robot Interaction* (pp. 443-451)
- Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making*, 5(5), 411-4
- Parasuraman, R., & Riley, V. (1997). Humans and Automation: Use, misuse, disuse, abuse. *Human Factors*, 39(2), 230-253. doi: 10.1518/001872097778543886

-
- Prahl, A., & Van Swol, L. (2017). Understanding algorithm aversion: When is advice from automation discounted? *Journal of Forecasting*, 36(6), 691-702. doi: 10.1002/for.2464
- Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J.-F., Breazeal, C., . . . Wellman, M. (2019). Machine behaviour. *Nature*, 568(7753), 477-486. doi: 10.1038/s41586-019-1138-y
- Rammert, W. (2011) Distributed Agency and Advanced Technology Or: How to Analyse Constellations of Collective Inter-Agency. *Technical University Technology Studies Working Papers*. Berlin.
- Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people*: Cambridge University Press: Cambridge, UK.
- Schroeder, J., & Epley, N. (2016). Mistaking Minds and Machines: How Speech Affects Dehumanization and Anthropomorphism. *Journal of Experimental Psychology-General*, 145(11), 1427-1437. doi: 10.1037/xge0000214
- Seeber, I., Bittner, E., Briggs, R. O., de Vreede, T., de Vreede, G.-J., Elkins, A., . . . Söllner, M. (2020). Machines as teammates: A research agenda on AI in team collaboration. *Information & Management*, 57(2), 103174. doi: <https://doi.org/10.1016/j.im.2019.103174>
- Seeber, I., Waizenegger, L., Seidel, S., Morana, S., Benbasat, I., & Lowry Paul, B. (2020). Collaborating with technology-based autonomous agents: Issues and research opportunities. *Internet Research*, 30(1), 1-18. doi: 10.1108/INTR-12-2019-0503
- Simmler, M., & Frischknecht, R. (2021). A taxonomy of human-machine collaboration: capturing automation and technical autonomy. *Ai & Society*, 36(1), 239-250. doi: 10.1007/s00146-020-01004-z
- Stafford, R. Q., MacDonald, B. A., Jayawardena, C., Wegner, D. M., & Broadbent, E. (2014). Does the Robot Have a Mind? Mind Perception and Attitudes Towards Robots Predict Use of an Eldercare Robot. *International Journal of Social Robotics*, 6(1), 17-32. doi: 10.1007/s12369-013-0186-y
- Sundar, S. S. (2020). Rise of Machine Agency: A Framework for Studying the Psychology of Human-AI Interaction (HAI). *Journal of Computer-Mediated Communication*, 25(1), 74-88. doi: 10.1093/jcmc/zmz026
- van Esch, P., Black, J. S., & Ferolie, J. (2019). Marketing AI recruitment: The next phase in job application and selection. *Computers in Human Behavior*, 90, 215-222. doi: 10.1016/j.chb.2018.09.009
- Wang, X. J., & Krumhuber, E. G. (2018). Mind Perception of Robots Varies With Their Economic Versus Social Function. *Frontiers in Psychology*, 9. doi: 10.3389/fpsyg.2018.01230

- 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. doi: 10.1177/1745691610369336
- Waytz, A., Gray, K., Epley, N., & Wegner, D. M. (2010). Causes and consequences of mind perception. *Trends in Cognitive Sciences*, 14(8), 383-388. doi: 10.1016/j.tics.2010.05.006
- 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. doi: 10.1016/j.jesp.2014.01.005
- Zlotowski, J., Proudfoot, D., Yogeeswaran, K., & Bartneck, C. (2015). Anthropomorphism: Opportunities and Challenges in Human-Robot Interaction. *International Journal of Social Robotics*, 7(3), 347-360. doi: 10.1007/s12369-014-0267-6

Article III – Technical autonomy in human-machine collaboration: Effects on the acceptance of advice and delegation of decisions

Ruth Frischknecht

Abstract

A growing number of tasks are being fulfilled through close collaboration between humans and increasingly autonomous machines. Successful deployment of such autonomous technologies depends heavily on how people (i.e., their potential co-workers) respond to them. However, little is known regarding how humans respond to technical systems with varying levels of autonomy. A vignette-based online experiment (N = 341) comprised of two decision scenarios was conducted to test the effects of different levels of technical autonomy on individuals' willingness to accept advice from and delegate decisions to such systems. The findings revealed that the level of technical autonomy did matter to human respondents. Both the willingness to accept advice and delegate decisions were more pronounced if the technical system was operating with a high level of technical autonomy. This pattern persisted across both decision scenarios. These findings contribute to a more differentiated understanding of human responses to increasingly autonomous technologies in situations where humans and machines collaborate.

Keywords: technical autonomy, machine learning, decision-making, human-machine collaboration, artificial intelligence

1 Introduction

In the past decade, smart machines have become an integral part of the work world. Often, they are no longer seen as mere tools and instead have become a type of co-worker (Seeber, Bittner, et al., 2020) or decision-maker (Helberger, Araujo, & de Vreese, 2020). In the right environment, this can create a “human-machine symbiosis” (Jarrahi, 2018). Thus, humans and machines are moving towards forming relationships, progressing beyond mere interaction (de Visser et al., 2018). Human collaboration with increasingly intelligent machines has the potential to augment decision-making (Jarrahi, 2018; Wilson & Daugherty, 2018), foster innovation (Brynjolfsson & McAfee, 2012; Raisch & Krakowski, 2021), elevate socio-technical capital (Makarius, Mukherjee, Fox, & Fox, 2020), and raise business revenue (Wilson & Daugherty, 2018). This is why a broad range of organizations, both private and public, already routinely deploy algorithms (Wilson & Daugherty, 2018). For example, such algorithms can be used to predict the likelihood of recidivism in criminal offenders (Dressel & Farid, 2018), select stocks in which to invest (Niszczoła & Kaszas, 2020), detect skin cancer (Haenssle et al., 2018), and assist designers with concept innovation (Wilson & Daugherty, 2018).

For machines to assist with or completely fulfill such complex tasks, they must be equipped with advanced technical capabilities. Thus, they are evolving to become “cognitive” (Watson, 2017). This means that an increasing number of such systems draw upon artificial intelligence (Russell & Norvig, 2016; Watson, 2017). Most applications related to artificial intelligence are based on machine learning, which is a machine’s ability to learn from and improve through experience (Jordan & Mitchell, 2015). However, this ability renders such machines increasingly opaque and difficult to comprehend (Danaher et al., 2017), as well as increasingly autonomous (de Visser et al., 2018; Simmler & Frischknecht, 2021). Consequently, machines with this type of ability have been referred to as “black boxes” (Rahwan et al., 2019). Great potential lies in combining the power of humans and machines (Jarrahi, 2018; Raisch &

Krakowski, 2021), but the ultimate impact these increasingly autonomous machines will have depends to a large extent on how people react to them.

The existing body of research on this topic is comprised of two generally contradictory responses to machines. Some humans tend to blindly trust and not reconsider recommendations made by technical systems, a condition referred to as overreliance (Goddard, Roudsari, & Wyatt, 2012; Parasuraman & Riley, 1997) or algorithm appreciation (Logg et al., 2019). Conversely, others are reluctant to rely on technical systems or fail to sufficiently incorporate the advice of machines, a state referred to as disuse of automation (Parasuraman & Riley, 1997) or with regards to newer technical systems, algorithmic aversion (Dietvorst, Simmons, & Massey, 2015). Research on how people perceive and respond to increasingly autonomous machines is crucial, as both algorithmic aversion and appreciation present potential pitfalls. In generally avoiding increasingly smart technical systems, there is the risk of rejecting technologies that might be beneficial, for example by fostering innovation (Wilson & Daugherty, 2018) or promoting welfare (Hidalgo, Orghian, Albo-Canals, De Almeida, & Martin, 2021). Conversely, by blindly adopting such systems, we may fail to recognize possibly negative consequences these systems may initiate (Hidalgo et al., 2021; O'Neil, 2016). Thus, for these types of systems to be used appropriately, it is important not only to understand the systems themselves, but also human perception of them (Seeber, Waizenegger, et al., 2020; Sundar & Kim, 2019).

In sum, there seems to be no clear tendency to either rely on (Logg et al., 2019) or avoid (Dietvorst et al., 2015) algorithms. In the current discussion, algorithmic forecasters are usually compared to human forecasters (Dietvorst et al., 2015; Logg et al., 2019) and algorithmic decisions to human decisions (Hoddinghaus et al., 2021; Lee, 2018). However, there is no such thing as “the algorithm” or the “recommender system.” Rather, algorithms vary in complexity and transparency, and these technical properties can influence the extent to which recommendations or advice from such systems and the systems themselves are accepted (Dietvorst et al., 2015). Furthermore, initial

expectations and assumptions about the algorithm (or technical system) and how it functions guide how the outputs generated by it are used (Burton, Stein, & Jensen, 2020). Applied to the context at hand, this suggests that it might not only matter *that* a technical system is entrusted with a particular task, but also *what* technical system is fulfilling it.

This research takes this development into account and explores how people respond to increasingly autonomous machines in decision-making situations. Thus, this study answers the following research question: *What is the effect of technical autonomy on the willingness to accept advice from and delegate decisions to a technical system?* In so doing, human responses to increasingly autonomous technical systems (in the form of a willingness to accept advice and delegate decisions) were tested in an online experiment. Specifically, the effect of technical autonomy was tested in two different organizational decision-making scenarios.

2 Theoretical Background

2.1 Algorithm aversion and algorithm appreciation

In many cases and across diverse domains, statistical models (i.e., algorithms) outperform human experts (Dawes, Faust, & Meehl, 1989; Grove, Zald, Lebow, Snitz, & Nelson, 2000). As noted above, despite their relative superiority in many application domains, people are averse to using algorithms, a condition referred to as algorithm aversion (Dietvorst et al., 2015). Likewise, Longoni, Bonezzi, and Morewedge (2019) found that consumers (here, patients) were unwilling to use medical artificial intelligence, even if the technical system clearly outperformed human doctors. People seem to be especially sensitive to technical systems making mistakes (Dietvorst et al., 2015; Prah & Van Swol, 2017). It has been argued that from the perspective of human operators, algorithms must be perfect to be used for decision-making, whereas humans are allowed to be imperfect (Burton et al., 2020; Dietvorst & Bharti, 2020). This

suggests that people apply stricter success criteria to the performance of an algorithm than they do to that of a human (Dietvorst & Bharti, 2020). Algorithmic aversion has not only been witnessed in forecasting tasks, but also in cooperation. In one experiment, people who believed they were cooperating with a bot in a prisoner's dilemma game transferred less money than they did when they believed they were cooperating with a human (Ishowo-Oloko et al., 2019).

However, there does not seem to be a general overall aversion towards algorithms (Prahl & Van Swol, 2017). Under certain conditions, machine outputs are regarded as more objective and rational than when humans accomplish the same tasks (Araujo et al., 2020; Dijkstra, Liebrand, & Timminga, 1998). Especially when there is no direct performance feedback (e.g., no erroneous predictions), people are more likely to rely on algorithmic advice across a wide array of forecasting tasks, a condition referred to as algorithm appreciation (Logg et al., 2019). Additionally, if people are able to adjust the forecast (or predictions) produced by an algorithm, they are more likely to rely on it (Dietvorst, Simmons, & Massey, 2018). An empirical study testing ideal partnerships between humans and machines found that managers were willing to accept machine input as long as it was weighted less than their own judgement (Haesevoets, De Cremer, Dierckx, & Van Hiel, 2021). Moreover, providing information regarding how an algorithm works has been found to increase its use. Participants who were informed about how an algorithm detecting skin cancer processed input data (i.e., medical information) to generate the output (i.e., the diagnosis) were less reluctant to use it than they were when they received no description of the process (Cadario, Longoni, & Morewedge, 2021).

Whether or not an algorithm is rejected seems not to depend solely on its status as a technical agent. Rather, the task that a technical system fulfills also seems to play a role. Especially in tasks that are considered objective, quantitative, and measurable (such as stock market and weather forecasts), the level of trust is high, and can be even higher than when a human fulfils the same task (Castelo,

Bos, & Lehmann, 2019). This pattern was emphasized by Lee (2018), who found that people perceive decisions made by an algorithm as equally fair and trustworthy when the task requires mechanical skill (such as scheduling shifts). In contrast, when a task is regarded as requiring human skills (such as evaluating performance), human decisions are perceived as fairer and more trustworthy. One study found that attitudes toward algorithmic decision-making are either equally (in the domain of media) or more favorable (in the domains of health care and justice) than attitudes about human decision-making (Araujo et al., 2020).

2.2 Autonomous technologies

As explained above, current technical developments are diverse and include a multitude of algorithms and systems that operate and function differently from one another (Araujo et al., 2020). To capture this development properly and understand the many facets of advanced technologies, it is necessary to conceive of technical autonomy as gradual and not binary (Endsley, 2017; Simmler & Frischknecht, 2021). In other words, different technical systems operate at different levels of autonomy and are not simply autonomous or non-autonomous. In their taxonomy, Simmler and Frischknecht (2021) described technical autonomy as comprised of four dimensions: transparency, indeterminism, adaptability, and openness. Based on that formulation, they argued that there are five consecutive levels of autonomy, suggesting that technical autonomy is not binary but rather a gradual property. Every level in the taxonomy is defined by a specific combination of dimensions, which are considered either present or absent. For example, a technical system operating on the highest level of autonomy is said to be non-transparent, indetermined, adaptable, and open. Likewise, a technical system operating on the lowest level of autonomy would be described as transparent, determined, non-adaptable, and closed. Thus, this taxonomy not only acknowledges the increasing amount of technical autonomy, but also categorizes and operationalizes different levels of autonomy.

Increasing levels of technical autonomy have been found to alter perceptions of competence and agency. Participants ascribed more agency (i.e., the ability to plan and act) and competence to technical systems high in technical autonomy, as compared to systems encompassing lower levels of technical autonomy (Frischknecht, 2021). This suggests that the level of technical autonomy alters user perceptions and thus may be a dimension relevant to shaping how outputs of algorithms are perceived. Machine intelligence (which entails capabilities such as learning or interactivity) is discussed as a significant driver for trust (Glikson & Woolley, 2020). That a technical system's level of autonomy matters is further demonstrated by Berger, Adam, Ruhr, and Benlian (2021), who showed that people relied more on the advice of algorithms capable of learning than they did on algorithms that were strictly determined. This suggests that the autonomy level may play a role in whether or not algorithmic advice is accepted. Existing research has not yet systematically incorporated the level of technical autonomy on which a system (or algorithm) operates. Thus, existing research has not sufficiently accounted for such technological diversity.

2.3 Hypotheses Development

As explained above, the goals of this study included investigating whether technical autonomy has an effect on human willingness to accept a technical system's advice and delegate decisions to it and determine how best to describe this effect. If people are to accept advice given by an algorithmic system, it is important that they understand how the output is produced (Cadario et al., 2021) and consider the technology comprehensible (Yeomans, Shah, Mullainathan, & Kleinberg, 2019). Closely related to the acceptance of advice is the question of whether one would delegate decisions to a machine (in other words, if one would let the machine decide). For humans to delegate decisions to a technology would mean depending on it beyond simply taking its advice (Komiak & Benbasat, 2006). Longoni et al. (2019) found that people did not want to use a medical decision system, especially if it replaced a human doctor. If the system supported a doctor in their work, this reluctance disappeared.

Therefore, delegation represents an even more extreme standpoint and shifts the focus from human-machine collaboration (i.e., solving problems jointly) to more automated forms of task completion.

As argued previously, increasing levels of technical autonomy render technical systems more and more intelligent and capable of solving complex problems (see e. g. Wilson & Daugherty, 2018). This, however, comes at the cost of reducing understandability, as technical autonomy makes systems less transparent (Simmler & Frischknecht, 2021) or opaque (Rahwan et al., 2019). Transparency is an important factor in accepting recommendations and building trust (Shin, 2020; Shin & Park, 2019). From this, it follows that people will be reluctant to use technical systems that have technical autonomy. Accordingly, this reluctance will be particularly pronounced in situations where the technical system has a high amount of autonomy. Thus, the following hypotheses were derived:

H1a: The level of technical autonomy affects human willingness to accept advice from technical systems.

H1b: The more autonomous a technical system is, the less willing humans are to accept its recommendations.

H2a: The level of technical autonomy affects human willingness to delegate decisions to a technical system.

H2b: The more autonomous a technical system is, the less willing humans are to delegate decisions to it.

Human-machine collaboration is always context-dependent, and existing research has clearly indicated that acceptance and willingness to delegate differ based on the task the system fulfills (see e.g. Castelo et al., 2019; de Winter & Hancock, 2015). Thus, it was assumed that the willingness to accept advice and delegate decisions would be higher if the technical system is assisting in an investment task (i.e., a calculable, number-based task) than if the same system

is assisting in a personnel selection task (i.e., a social task requiring judgment and intuition). This led to the following hypotheses:

H3a: The willingness to accept the recommendation of a technical system is higher in the task domain of money investment than in the task domain of personnel selection.

H3b: The willingness to delegate a decision to a technical system is higher in the task domain of money investment than in the task domain of personnel selection.

3 Method

A vignette-based online experiment was conducted to test the hypotheses derived above. The taxonomy of Simmler and Frischknecht (2021) was used to derive the levels of technical autonomy. The vignettes described a technical system solving two different tasks. Two different task domains were included so that more generalizable conclusions could be drawn, and a determination made regarding whether and to what extent technical autonomy had an effect. The experiment also included a condition where a human expert made a recommendation. The human expert condition was included as a control condition, anchoring the results and allowing for better interpretation.

3.1 Design and independent variables

To test the above hypotheses, a 2 (task: personnel vs. investment) $\times$ 6 (level of autonomy: deterministic system vs. non-transparent system vs. indeterminated system vs. adaptable system vs. open system vs. human) mixed design was used. The participants were randomly assigned to one of the levels of technical autonomy (or the human condition) between subjects and rated the system for use in two different task domains. Accordingly, task domains were repeated within subjects. The level of autonomy and task domain were manipulated using the vignettes describing the different decision scenarios. Given the context and goal of the study, the vignettes were based on two organizational decision

scenarios: one regarding personnel and one that was financial in nature. As this factor was manipulated within-subjects, every participant received both scenarios. The order the two scenarios were presented was randomized to counterbalance any potential order effects. In the real world, algorithms are currently assisting humans in making decisions in both domains addressed by the vignettes. For example, advanced technical systems are currently applied in a wide array of personnel selection tasks (Langer, König, & Busch; Wilson & Daugherty, 2018). Likewise, advanced technical systems are also widespread in forecasting financial developments and suggesting investment strategies (Wilson & Daugherty, 2018; Zhang, Pentina, & Fan). In this way, the scenarios reflected realistic situations.

The level of autonomy was manipulated in accordance with the dimensions of technical autonomy developed by Simmler and Frischknecht (2021). Depending on the condition a participant was assigned, he or she received a companion description of how the algorithm reached its recommendation. The deterministic system (i.e., the “no autonomy” condition) was described as strictly determined, transparent, non-adaptable, and closed. The open system (i.e., the representing the highest amount of technical autonomy) was described as indeterminated, non-transparent, adaptable, and open. Thus, the level of autonomy was manipulated in a between-subjects manner. For a detailed description of how the autonomy levels in the vignettes were varied in the experiment condition, see Table 11 in Appendix A. The procedure for deriving and describing the different levels was analogous to the procedure described in Frischknecht (2021). As few words as possible were exchanged in the vignette and the information was always presented in the same order.

The most important words were highlighted in bold. Figure 4 depicts a sample vignette for the open system and human control condition (the vignettes for both task domains and every level of technical autonomy are provided in Appendix A). The vignette consisted of two parts: descriptions of the decision-making scenario and technical system used in that scenario. The participants in the human expert condition received only the first part.

You are a manager in a medium-sized company. You have been advertising a job opening lately. Now **you need to decide who to hire**. The company you work for **uses a technical system/engages an external expert to assist you** in this decision. **The recommendation system/the expert suggests a suitable** candidate for the job. However, it is **up to you to decide** if you want to follow the recommendation, or if you prefer to follow your own preference.

You want to understand how the technical system and algorithm behind it work, so you read the manual. There, it says:

- The recommender system is **able to learn** from past decision preferences. For example, it **adjusts its recommendations** based on what option other users chose.
- The system is **open and connected to other technical systems or the internet**. It **can access many different data sources**. Thus, it **expands the data** upon which it bases its recommendations. It is not possible for you to verify entirely the data upon which the suggestions are based.
- The recommender system is **not completely predetermined**, which means **the same input data will not necessarily lead to the same recommendation**.
- How the system gets from the input data to the recommendation is **neither fully traceable nor fully transparent to you**.

Figure 4: Sample vignette showing the scenario and describing the open system.

3.2 Dependent variables

Willingness to accept a recommendation and delegate a decision served as dependent variables. Both measures were adjusted from Komiak and Benbasat (2006). Willingness to accept a recommendation was measured with three items (e.g., I am willing to use this technical system to help with my decisions about whom to hire/where to invest money). The Cronbach's α was .91 for the personnel selection scenario and .90 for the investment scenario. The willingness to delegate a decision was also measured by three items (e.g., I am willing to delegate to this technical system my decision about whom to hire/where to invest money). The Cronbach's α was .93 for the personnel selection scenario and .94 for the investment scenario. The items were adapted to the context of this study and that of the respective scenario. Because the scale provided by Komiak and Benbasat (2006) only consisted of two items, a self-developed third item (i.e., I would hand over similar decisions to this technical system) was added. All items were measured on a seven-point Likert scale, with responses ranging from strongly disagree to strongly agree.

3.3 Covariates and additional variables

General trust in technology has been found to have an impact on the use of technical systems and acceptance of their output (Araujo et al., 2020; Hoddinghaus et al., 2021). Thus, the disposition to trust technologies was included as a covariate and measured with the scale from Lankton, McKnight, and Tripp (2015) (Cronbach's $\alpha = .82$). Participants rated their trust in technology on the same seven-point Likert scale used for the dependent variables. Task familiarity was also included as a covariate because it has been shown that experts are especially reluctant to rely on algorithmic advice (Logg et al., 2019). Task familiarity was measured with three generic items established by Kent and Allen (1994), using a semantic differential with seven steps (e.g., unfamiliar/familiar). The ratings from both tasks were collapsed into one familiarity measure (Cronbach's $\alpha = .90$).

Manipulation and attention checks. The “perceived autonomy scale” initially employed by Frischknecht (2021) was used as a manipulation check and consisted of the items *autonomous, independent, self-sufficient, and self-reliant* (Cronbach’s $\alpha = .87$). For identifying inattentive and straight-lining participants, three attentions checks were distributed throughout the questionnaire. Following Abbey and Meloy (2017), these attention checks took the form of directed queries (sample item: “*In this row, click strongly agree*”). Participants who failed to click the correct answer were excluded from the sample.

3.4 Procedure

After providing general information about the study, the participants read one of two decision scenarios and were instructed to picture it as vividly as possible. Then they received questions about the technical system described in the scenario, which already contained the first set of dependent variables. This was followed by a description of a second scenario about which the participants had to answer the same questions, though slightly adjusted to fit the new information. In this fashion, the second set of dependent variables was measured. After completing the questions with regards to the dependent variables, the participants filled in their impressions concerning the covariates. Next, they answered questions about trusting technology and their familiarity with investment and personnel selection decisions. Then, they were instructed to provide demographic information such as age and amount of education. Lastly, the participants received a code for payment. All items for capturing the dependent variables and covariates and the manipulation and attention checks appeared in randomized order to counterbalance any potential effects of presentation.

3.5 Sample

A power analysis using G*Power was conducted to determine the sample size for the experiment. To detect a small interaction effect in a repeated measures analysis of variance (ANOVA) ($\alpha = 0.05$, $1 - \beta = 0.80$), 330 participants were required. The participants were recruited on M-Turk. Only workers with more than a 95% approval rate and living in the United States who had already completed more than 500 worker tasks were admitted to the online experiment. Overall, 420 participants completed the survey experiment (161 females, 256 males, and three participants who either indicated “other” for gender or did not answer this question; mean age = 39.88 years, $SD = 11.43$). Please note that the oversampling was intended, as it was assumed that participants might either fail an attention check or leave the study for an unspecified reason. Indeed, we had to exclude 79 participants because they failed at least one of the three attention checks. The final sample included 341 participants (123 females, 215 males, and three participants who either indicated “other” for gender or did not answer this question). The mean age of the participants in the final sample was 39.97 years ($SD = 11.53$).

4 Results

4.1 Preliminary analyses

To verify whether the vignettes successfully manipulated technical autonomy, a one-way ANOVA was conducted with the levels of autonomy serving as a between-subjects factor and perceived autonomy as the dependent variable. Because the assumption of homogeneity of variances was violated, a robust Welch ANOVA was run. According to that analysis, there was a statistically significant main effect of the level of autonomy on the amount of perceived autonomy ($F(4, 140) = 3.15$, $p = 0.016$, $\eta_p^2 = .041$), indicating that the perceived level of autonomy differed across the five different conditions. Thus, the manipulation of technical autonomy was considered successful.

Before testing the hypotheses, all assumptions for a two-way mixed ANOVA were tested. This initial test revealed that the data obtained in this study violated all relevant assumptions. The data were not distributed normally and there was no homogeneity of variances (Field, 2018). The data were evaluated with a regular mixed ANOVA and the results compared to a robust version via the WRS2 package in R (Mair & Wilcox, 2020). The comparison of the two analyses showed that the pattern of results and significance of effects were the same for both.¹ For better comprehensibility and readability, only the results of the regular mixed ANOVA are reported herein. Because both covariates – general trust in technology and task familiarity – did not alter the pattern of results, they were excluded from the final analysis. Descriptive statistics of all dependent variables is provided in Table 10.

¹ A robust ANOVA using the bwtrim command in R revealed a statistically significant main effect of the level of autonomy on the acceptance of advice ($F(4, 81) = 7.27, p < 0.001$) but no statistically significant main effect of task domain ($F(1, 131.55) = 0.79, p = .38$), as well as no statistically significant interaction ($F(4, 83.896) = 1.91, p = 0.12$). The robust ANOVA with a willingness to delegate decisions to the technical system serving as a dependent variable revealed a statistically significant main effect of technical autonomy ($F(4, 84.25) = 3.29, p = 0.015$) and not a main effect of task domain ($F(1, 155.83) = 1.17, p = .28$), as there was no interaction ($F(4, 84.05) = 1.07, p = .376$).

Table 10: Descriptive statistics of the dependent variables.

Measure	Deterministic System		Non-transparent System		Indetermined System		Adaptable System		Open System		Human Expert	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Personnel Decision												
Willingness to accept	4.55	1.67	4.53	1.72	4.24	1.56	5.03	1.37	5.3	1.20	5.73	0.90
Willingness to delegate	3.57	1.87	3.76	1.96	3.47	1.83	4.17	1.66	4.15	1.74	4.16	1.72
Investment Decision												
Willingness to accept	4.23	1.70	4.48	1.64	4.29	1.64	5.13	1.17	5.51	1.03	5.89	0.81
Willingness to delegate	3.52	2.01	3.73	1.92	3.42	1.88	4.36	1.70	4.48	1.54	4.64	1.57

4.2 Hypotheses testing

A two-way mixed ANOVA was conducted to investigate the impacts of the level of technical autonomy and task domain on advice acceptance. According to this analysis, there was a statistically significant main effect of the level of technical autonomy on the willingness to accept advice ($F(4, 280) = 7.29, p < 0.001, \eta_p^2 = .094$). This result supports Hypothesis 1a. Additionally, there was no statistically significant main effect of task domain ($F(1, 280) = 0.002, p = .96, \eta_p^2 < .001$) on the willingness to accept advice from the recommender system. This result is contrary to Hypothesis 3a. Moreover, there was no statistically significant interaction between the level of technical autonomy and task domain ($F(4, 280) = 1.904, p = 0.11, \eta_p^2 = .026$) for acceptance of the technical system as a decision aid. Post hoc tests using the Games-Howell adjustment (which is best to use when the variances are unequal) revealed that the participants were significantly more willing to accept advice from the open system than from the deterministic system ($p < .001, 1.01, 95\% \text{ CI } [.35 - 1.68]$), non-transparent system ($p = .005, .90, 95\% \text{ CI } [.20 - 1.61]$), or indetermined system ($p < .001, 1.14, 95\% \text{ CI } [.47 - 1.81]$). Likewise, recommendations from the adaptable system were regarded as more acceptable than recommendations from the indetermined system ($p = .015, .81, 95\% \text{ CI } [.11 - 1.51]$). Thus, Hypothesis 1b was not supported. All other groups did not differ statistically significantly from one another. Figure 5 depicts the group differences.

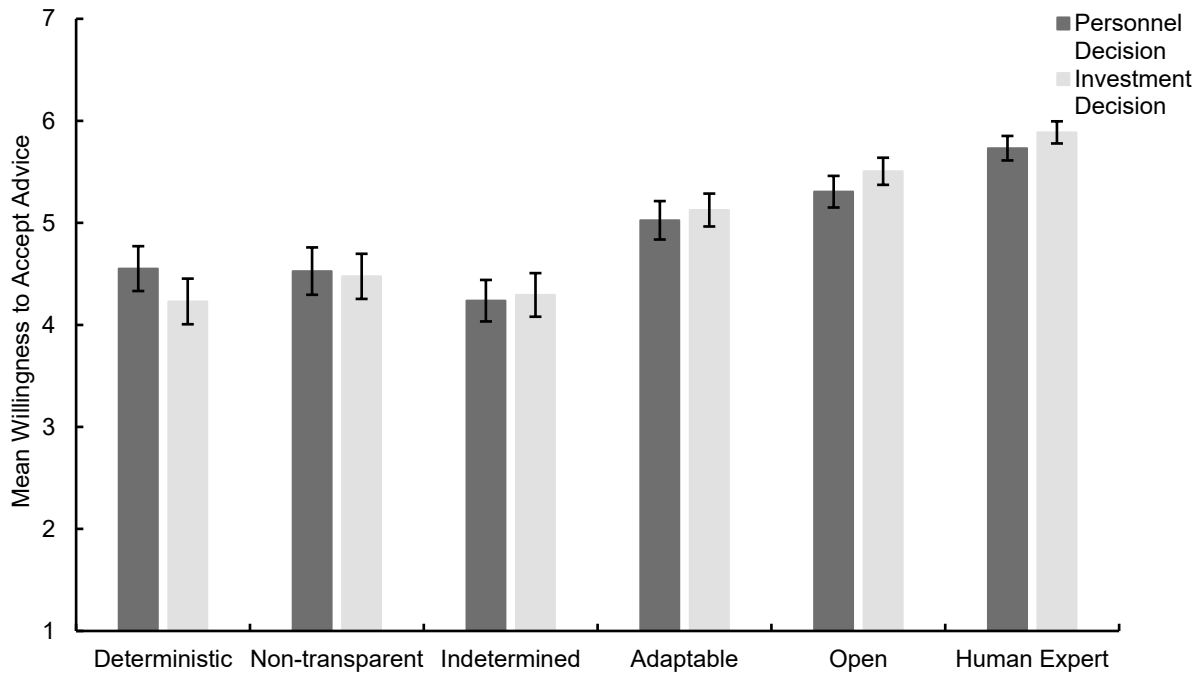


Figure 5: Mean willingness to accept advice from a recommender system in personnel and investment decisions across all experiment conditions. Error bars represent standard errors.

Another two-way mixed ANOVA was conducted to investigate the effects of the level of technical autonomy and task domain on willingness to delegate a decision to the technical system. According to this analysis, there was a statistically significant main effect of technical autonomy ($F(4, 380) = 3.11, p = 0.016$, partial $\eta_p^2 = .043$), indicating that there was a statistically significant difference in the means of willingness to delegate a decision across the different levels of technical autonomy. These results provide support for Hypothesis 2a. There was no statistically significant main effect of task domain, indicating that there were no differences in the willingness to delegate a decision among the task domains ($F(1, 280) = 1.66, p = .2, \eta_p^2 = .006$). This is contrary to Hypothesis 3b. There was no statistically significant interaction between technical autonomy and task domain with regards to willingness to delegate, ($F(4, 280) = 1.71, p = .149, \eta_p^2 = .024$).

Counter to Hypothesis 2b, post hoc tests using the Games-Howell adjustment revealed that people were more willing to delegate a decision to the open system than to the indetermined system ($p = .043, .87, 95\% \text{ CI } [.02 - 1.73]$). Figure 6 depicts the group differences.

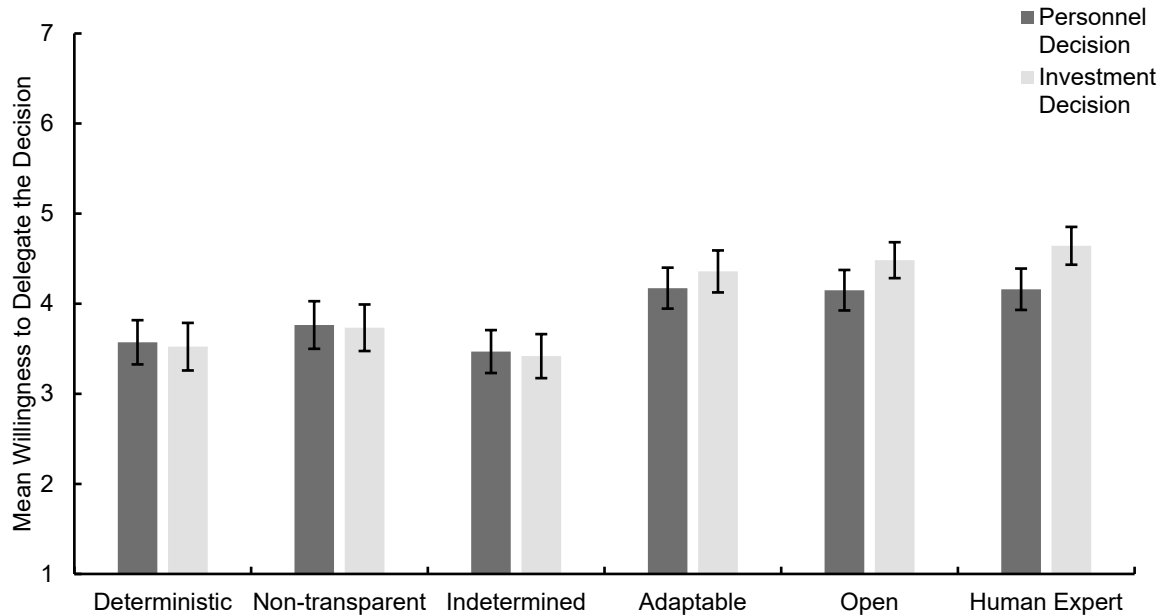


Figure 6: Mean willingness to delegate a decision to the recommender system across the different decision scenarios and six experiment conditions. Error bars represent standard errors.

4.3 Additional analyses

Additionally, the same analyses were run again, this time including the human expert condition. The pattern of results persisted. A two-way mixed ANOVA showed that there was a statistically significant main effect of the level of autonomy on acceptance ($F(5, 93.55) = 10.25, p < .001$). Neither the main effect of task domain ($F(1, 151.41) = 0.3, p = 0.59$) nor the interaction between technical autonomy and task domain ($F(5, 95.05) = 1.63, p = .16$) were found to be statistically significant. Post-hoc comparisons using the Games-Howell adjustment revealed that the participants were statistically significantly more willing to accept advice from a human expert than from the deterministic system ($p < .001, 1.42, 95\% \text{ CI} [.76 - 2.08]$), non-transparent system ($p < .001, 1.31, 95\% \text{ CI} [.61 - 2.00]$), indetermined system ($p < .001, 1.54, 95\% \text{ CI} [.88 - 2.21]$), or adaptable system ($p = .002, .73, 95\% \text{ CI} [.20 - 1.26]$). However, ratings between the open system and human expert did not differ statistically significantly.

Another two-way mixed ANOVA was conducted that revealed statistically significant main effects of autonomy ($F(5, 95.37) = 3.37, p = .007$) and task domain

($F(1, 184.8) = 4.58, p = .033$) on the willingness to delegate the decision. The interaction of autonomy and task domain was found to be non-significant ($F(5, 95.4) = 1.62, p = 0.16$). Another set of post hoc tests using the Games-Howell adjustment showed that people were more willing to delegate a decision to a human expert than an indetermined system ($p = .028, .96, 95\% \text{ CI } [.06 - 1.85]$). All other groups did not differ statistically significantly from one another (smallest $p = .06$).

5 Discussion

Joint decision-making between humans and machines has become increasingly widespread (Seeber, Waizenegger, et al., 2020) and the machines involved in these decisions increasingly autonomous (Simmler & Frischknecht, 2021). This study explored the effects of different levels of technical autonomy on the willingness to accept advice from and delegate decisions to such systems. Overall, the findings suggest that increased technical autonomy matters for both. This indicates that there is neither a general aversion to (Dietvorst et al., 2015) nor general appreciation (Logg et al., 2019) for such technical systems. Rather, willingness to use them (either by accepting their advice or allowing them to choose) seems to depend on their technical function. Furthermore, this means that not only human characteristics such as expertise (Burton et al., 2020; Logg et al., 2019) or level of trust in technology (Hoddinghaus et al., 2021), but also the technical capabilities (or characteristics) of the system itself play a role in the joint decision-making processes.

In addition, the results show that the influence of technical autonomy persists across different task domains. Thus, the decision context does not seem to matter; people are equally willing or unwilling to rely on the particular technical system. This finding contradicts the hypotheses posed in this study and the existing body of research, which states that technical systems seem to be accepted mainly for quantitative objective tasks, but not for social tasks requiring intuition (de Winter & Hancock, 2015; Sundar & Kim, 2019). Taken

together, these general findings provide initial evidence of the relevance of technical autonomy in human machine collaboration.

More specifically, the results show that the willingness to accept advice and delegate decisions is more pronounced for technical systems operating with high rather than lower levels of technical autonomy. This contradicts the hypotheses developed for this experiment and many theoretical considerations positing that people would be reluctant to use opaque technical systems and instead prefer explainable and transparent systems (Cadario et al., 2021; Glikson & Woolley, 2020; Yeomans et al., 2019). Rather, these results suggest that participants valued technical capabilities (or capacities) more than they did comprehensibility, traceability, or transparency.

The increase in willingness to accept advice and delegate decisions became especially evident when the system was described as capable of adapting and learning. This complements the findings of Berger et al. (2021) and the proposition of Glikson and Woolley (2020) by further highlighting that the ability to learn may be an important factor in the acceptance of and adherence to advice given by algorithms and technical systems. Both the present work and Berger et al. (2021) found that it may be this ability to learn that convinces users to rely on algorithms. This tendency seems to be even more pronounced if the system is open and thus capable of interacting. Past research has also shown that technical systems operating with high levels of autonomy are ascribed more competence and agency than are technical systems operating with less (Frischknecht, 2021). This underscores that the importance of technical autonomy is particularly manifested where systems (or algorithms) are adaptable and open, thus creating a shift in social ascription and the willingness to use such systems. Interestingly, a comparison of the different technical autonomy conditions and human expert condition showed that ratings related to the open system did not differ from those of the human expert. Thus, advice coming from a technical system operating on the highest level of technical autonomy captured in the Simmler and Frischknecht (2021) seems to be of equal

consideration as that of a human expert. This not only applies to the willingness to accept advice but also for delegating decisions.

The outcomes presented here have a number of practical implications. As the results suggest that people tend to accept advice from and delegate decisions to technical systems that are described as adaptable and open, emphasizing these particular technical properties should foster adherence to the advice from and adoption of such systems. Thus, if an organization or manager seeks to foster the use of a given technical system that is, for example, based on a machine learning algorithm, highlighting this technical property might be useful. Conversely, the results also indicate that people were quite willing to accept advice from technical systems high in autonomy. Outputs from such systems are difficult to trace and hard to explain. Measures such as comprehensive training and information campaigns should therefore be implemented before technical systems with a high degree of autonomy are employed. At the same time, it would be useful to enforce certain precautionary measures in the design of such systems, so that users do not begin blindly accepting advice given by increasingly autonomous systems.

Despite the clear pattern, the findings presented here should be interpreted with caution. Firstly, the data were gathered in an online experiment. Therefore, the results have little validity regarding the influence of different levels of technical autonomy in decision-making in real-world settings. This is especially the case because situations outside the laboratory are complex and guided by very different and vague interests. The effect of technical autonomy in more realistic conditions is likely to be substantially smaller. Secondly, the manipulation of different levels of technical autonomy was based on vignettes that strongly emphasized certain technical capabilities. What a technical system is capable of, however, is usually not as obvious as it was in this study. Additionally, (Cadario et al., 2021) demonstrated that people were more willing to use an algorithmic health provider if it was explained how the algorithm worked. However, the amount of information about technical functioning provided in the vignettes was kept equal for all participants rating the technical system. Thus, the amount

of information about how the technical system functioned is an unlikely explanation for the pattern of findings.

Thirdly, this study captures the willingness to accept advice from and delegate decisions to technical systems operating on different levels of autonomy. Thus, the dependent variable measured an intention and not actual behavior. What influence increasing technical autonomy has on real behavior (e.g., on the actual decisions made) remains to be seen. Future research should therefore use additional behavioral measures. Likewise, the vignettes did not contain actual advice, and rather only described a decision scenario.

Despite these limitations and further questions raised by this study, it was among the first to test the fundamental influence of increasing technical autonomy on decision-making situations. As this study revealed that increasing technical autonomy plays a role in collaborative decision-making, future research should systematically investigate this variable. There are also several other promising areas for future research. As implied above, there is great potential for real-world investigations exploring how people respond to systems that employ high levels of autonomy, how individuals organize themselves around increasingly autonomous systems, and how they make sense of the output of such machines. Future research could also examine how decisions change and suggestions from increasingly autonomous machines are weighted and incorporated into individual and collective decisions. Another promising area for further research is an individual's expertise or professional background. Previous work has shown that experts are more reluctant to use algorithmic advice than are lay persons (Logg et al., 2019); thus, scholars could further investigate if and how this reluctance applies to increasing levels of technical autonomy. Especially in professional organizations, decisions are made by experts. Thus, understanding how such experts (e.g., judges, bankers, doctors) respond to increasingly autonomous decision support systems would be a relevant research endeavor.

6 Conclusion

Because of the ongoing spread of increasingly autonomous machines and their deployment in a variety of organizations, situations in which humans and machines closely collaborate have become more prevalent. Human responses to such increasingly autonomous machines are not well-understood. Most research on this topic has compared responses to machines with responses to humans (Dietvorst et al., 2015; Hoddinghaus et al., 2021; Lee, 2018; Logg et al., 2019) and have not systematically varied technical functioning nor differentiated among different levels of technical autonomy. However, current technical development allows for technical systems (or algorithms) to operate on various levels of technical autonomy. The evidence presented in this research indicates that human responses change depending on the level of technical autonomy on which a system operates. More precisely, under the condition of increasing autonomy, the willingness to accept advice from and delegate decisions to technical systems grows. For human users, there does not seem to be such a thing as "the algorithm" or "the system." Rather, such machines tend to be differentiated according to how the particular technology operates. Thus, technical autonomy with its various levels is an important variable that requires careful consideration in human-machine collaboration.

Appendix A

Table 11: Technical system descriptions provided in the vignettes, organized according to the level of autonomy.

Deterministic System	The recommender system is not able to learn from past decision preferences and it does not adjust its recommendations based on what option previous users chose. The system is closed and not connected to other technical systems or the internet. The data upon which the system bases its recommendations is specified in advance. This means that the data source is predetermined and cannot be extended. The recommender system is completely predetermined, which means the same input always leads to the same recommendation. How the system gets from the input data to the recommendation is fully traceable and transparent to you.
Non-transparent System	The recommender system is not able to learn from past decision preferences and it does not adjust its recommendations based on what option previous users chose. The system is closed and not connected to other technical systems or the internet. The data upon which the system bases its recommendations is specified in advance. This means that the data source is predetermined and cannot be extended. The recommender system is completely predetermined, which means the same input always leads to the same recommendation. How the system gets from the input data to the recommendation is neither fully traceable nor fully transparent to you.
Indetermined System	The recommender system is not able to learn from past decision preferences and it does not adjust its recommendations based on what option previous users chose. The system is closed and not connected to other technical systems or the internet. The data upon which the system bases its recommendations is specified in advance. This means that the data source is predetermined and cannot be extended. The recommender system is not completely predetermined, which means the same input data will not necessarily lead to the same recommendation. How the system gets from the input data to the recommendation is neither fully traceable nor fully transparent to you.
Adaptable System	The recommender system is able to learn from past decision preferences. For example, it adjusts its recommendations based on what option other users chose. The system is closed and not connected to other technical systems or the internet. The data upon which the system bases its recommendations is specified in advance. This means that the data source is predetermined and cannot be extended. The recommender system is not completely predetermined, which means the same input data will not necessarily lead to the same recommendation. How the system gets from the input data to the recommendation is neither fully traceable nor fully transparent to you.
Open System	The recommender system is able to learn from past decision preferences. For example, it adjusts its recommendations based on what option other users chose. The system is open and connected to other technical systems or the internet. It can access many different data sources. Thus, it expands the data upon which it bases its recommendations. The recommender system is not completely predetermined, which means the same input data will not necessarily lead to the same recommendation. How the system gets from the input data to the recommendation is neither fully traceable nor fully transparent to you.

7 References

- Abbey, J. D., & Meloy, M. G. (2017). Attention by design: Using attention checks to detect inattentive respondents and improve data quality. *Journal of Operations Management*, 53-56(1), 63-70. doi: 10.1016/j.jom.2017.06.001
- Araujo, T., Helberger, N., Kruikemeier, S., & de Vreese, C. H. (2020). In AI we trust? Perceptions about automated decision-making by artificial intelligence. *Ai & Society*, 35(3), 611-623. doi: 10.1007/s00146-019-00931-w
- Berger, B., Adam, M., Ruhr, A., & Benlian, A. (2021). Watch Me Improve-Algorithm Aversion and Demonstrating the Ability to Learn. *Business & Information Systems Engineering*, 63(1), 55-68. doi: 10.1007/s12599-020-00678-5
- Brynjolfsson, E., & McAfee, A. (2012). Winning the Race With Ever-Smarter Machines. *Mit Sloan Management Review*, 53(2), 53-60.
- Burton, J. W., Stein, M. K., & Jensen, T. B. (2020). A systematic review of algorithm aversion in augmented decision making. *Journal of Behavioral Decision Making*, 33(2), 220-239. doi: 10.1002/bdm.2155
- Cadario, R., Longoni, C., & Morewedge, C. K. (2021). Understanding, explaining, and utilizing medical artificial intelligence. *Nature Human Behaviour*. doi: 10.1038/s41562-021-01146-0
- Castelo, N., Bos, M. W., & Lehmann, D. R. (2019). Task-Dependent Algorithm Aversion. *Journal of Marketing Research*, 56(5), 809-825. doi: 10.1177/0022243719851788
- Danaher, J., Hogan, M. J., Noone, C., Kennedy, R., Behan, A., De Paor, A., . . . Shankar, K. (2017). Algorithmic governance: Developing a research agenda through the power of collective intelligence. *Big Data & Society*, 4(2), 2053951717726554. doi: 10.1177/2053951717726554
- Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science*, 243(4899), 1668-1674. doi: 10.1126/science.2648573
- de Visser, E. J., Pak, R., & Shaw, T. H. (2018). From "automation" to "autonomy": the importance of trust repair in human-machine interaction. *Ergonomics*, 61(10), 1409-1427. doi: 10.1080/00140139.2018.1457725
- de Winter, J. C. F., & Hancock, P. A. (2015). Reflections on the 1951 Fitts List: Do Humans Believe Now that Machines Surpass them? *Procedia Manufacturing*, 3, 5334-5341. doi: 10.1016/j.promfg.2015.07.641
- Dietvorst, B. J., & Bharti, S. (2020). People Reject Algorithms in Uncertain Decision Domains Because They Have Diminishing Sensitivity to Forecasting Error. *Psychological Science*, 31(10), 1302-1314. doi: 10.1177/0956797620948841

-
- 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. doi: 10.1037/xge0000033
- Dietvorst, B. J., Simmons, J. P., & Massey, C. (2018). Overcoming Algorithm Aversion: People Will Use Imperfect Algorithms If They Can (Even Slightly) Modify Them. *Management Science*, 64(3), 1155-1170. doi: 10.1287/mnsc.2016.2643
- Dijkstra, J. J., Liebrand, W. B. G., & Timminga, E. (1998). Persuasiveness of expert systems. *Behaviour & Information Technology*, 17(3), 155-163. doi: 10.1080/014492998119526
- Dressel, J., & Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism. *Science Advances*, 4(1). doi: 10.1126/sciadv.aao5580
- Endsley, M. R. (2017). From Here to Autonomy: Lessons Learned From Human-Automation Research. *Human Factors*, 59(1), 5-27. doi: 10.1177/0018720816681350
- Field, A. (2018). *Discovering statistics using IBM SPSS statistics*. Los Angeles: SAGE.
- Frischknecht, R. (2021). A social cognition perspective on autonomous technology. *Computers in Human Behavior*, 122, 106815. doi: <https://doi.org/10.1016/j.chb.2021.106815>
- Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical evidence. *Academy of Management Annals*, 14(2), 627-660. doi: 10.5465/annals.2018.0057
- Goddard, K., Roudsari, A., & Wyatt, J. C. (2012). Automation bias: a systematic review of frequency, effect mediators, and mitigators. *Journal of the American Medical Informatics Association*, 19(1), 121-127. doi: 10.1136/amiajnl-2011-000089
- Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment*, 12(1), 19-30. doi: 10.1037/1040-3590.12.1.19
- Haenssle, H. A., Fink, C., Schneiderbauer, R., Toberer, F., Buhl, T., Blum, A., . . . Zalaudek, I. (2018). Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Ann Oncol*, 29(8), 1836-1842. doi: 10.1093/annonc/mdy166
- Haesevoets, T., De Cremer, D., Dierckx, K., & Van Hiel, A. (2021). Human-machine collaboration in managerial decision making. *Computers in Human Behavior*, 119. doi: 10.1016/j.chb.2021.106730
- Helberger, N., Araujo, T., & de Vreese, C. H. (2020). Who is the fairest of them all? Public attitudes and expectations regarding automated decision-

-
- making. *Computer Law & Security Review*, 39. doi: 10.1016/j.clsr.2020.105456
- Hidalgo, C., Orghian, D., Albo-Canals, J., De Almeida, F., & Martin, N. (2021). *How humans judge machines*. Cambridge, Massachusetts: MIT Press.
- Hoddinghaus, M., Sondern, D., & Hertel, G. (2021). The automation of leadership functions: Would people trust decision algorithms? *Computers in Human Behavior*, 116. doi: 10.1016/j.chb.2020.106635
- Ishowo-Oloko, F., Bonnefon, J. F., Soroye, Z., Crandall, J., Rahwan, I., & Rahwan, T. (2019). Behavioural evidence for a transparency-efficiency tradeoff in human-machine cooperation. *Nature Machine Intelligence*, 1(11), 517-+. doi: 10.1038/s42256-019-0113-5
- Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business Horizons*, 61(4), 577-586. doi: 10.1016/j.bushor.2018.03.007
- Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255-260. doi: 10.1126/science.aaa8415
- Kent, R. J., & Allen, C. T. (1994). Competitive Interference Effects in Consumer Memory for Advertising: The Role of Brand Familiarity. *Journal of Marketing*, 58(3), 97-105. doi: 10.2307/1252313
- Komiak, S. Y. X., & Benbasat, I. (2006). The Effects of Personalization and Familiarity on Trust and Adoption of Recommendation Agents. *MIS Quarterly*, 30(4), 941-960. doi: 10.2307/25148760
- Langer, M., Konig, C. J., & Busch, V. Changing the means of managerial work: effects of automated decision support systems on personnel selection tasks. *Journal of Business and Psychology*. doi: 10.1007/s10869-020-09711-6
- Lankton, N. K., 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.
- Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*, 5(1), 2053951718756684. doi: 10.1177/2053951718756684
- 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. doi: 10.1016/j.obhdp.2018.12.005
- Longoni, C., Bonezzi, A., & Morewedge, C. K. (2019). Resistance to Medical Artificial Intelligence. *Journal of Consumer Research*, 46(4), 629-650. doi: 10.1093/jcr/ucz013
- Mair, P., & Wilcox, R. (2020). Robust statistical methods in R using the WRS2 package. *Behavior Research Methods*, 52(2), 464-488. doi: 10.3758/s13428-019-01246-w

-
- Makarius, E. E., Mukherjee, D., Fox, J. D., & Fox, A. K. (2020). Rising with the machines: A sociotechnical framework for bringing artificial intelligence into the organization. *Journal of Business Research*, 120, 262-273. doi: 10.1016/j.jbusres.2020.07.045
- Niszczoła, P., & Kaszas, D. (2020). Robo-investment aversion. *Plos One*, 15(9). doi: 10.1371/journal.pone.0239277
- O'Neil, C. (2016). *Weapons of math destruction : how big data increases inequality and threatens democracy*. New York: Crown.
- Parasuraman, R., & Riley, V. (1997). Humans and Automation: Use, misuse, disuse, abuse. *Human Factors*, 39(2), 230-253. doi: 10.1518/001872097778543886
- Prahl, A., & Van Swol, L. (2017). Understanding algorithm aversion: When is advice from automation discounted? *Journal of Forecasting*, 36(6), 691-702. doi: 10.1002/for.2464
- Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J.-F., Breazeal, C., . . . Wellman, M. (2019). Machine behaviour. *Nature*, 568(7753), 477-486. doi: 10.1038/s41586-019-1138-y
- Raisch, S., & Krakowski, S. (2021). Artificial Intelligence and Management: The Automation-Augmentation Paradox. *The Academy of Management Review*, 46(1), 192. doi: 10.5465/amr.2018.0072
- Russell, S. J., & Norvig, P. (2016). *Artificial intelligence : a modern approach*. Boston: Pearson Education.
- Seeber, I., Bittner, E., Briggs, R. O., de Vreede, T., de Vreede, G.-J., Elkins, A., . . . Söllner, M. (2020). Machines as teammates: A research agenda on AI in team collaboration. *Information & Management*, 57(2), 103174. doi: <https://doi.org/10.1016/j.im.2019.103174>
- Seeber, I., Waizenegger, L., Seidel, S., Morana, S., Benbasat, I., & Lowry Paul, B. (2020). Collaborating with technology-based autonomous agents: Issues and research opportunities. *Internet Research*, 30(1), 1-18. doi: 10.1108/INTR-12-2019-0503
- Simmler, M., & Frischknecht, R. (2021). A taxonomy of human-machine collaboration: capturing automation and technical autonomy. *Ai & Society*, 36(1), 239-250. doi: 10.1007/s00146-020-01004-z
- Sundar, S. S., & Kim, J. (2019). Machine heuristic: When we trust computers more than humans with our personal information. In: *Proceedings of the 2019 CHI Conference on human factors in computing systems* (pp. 1-9).
- Watson, H. J. (2017). Preparing for the Cognitive Generation of Decision Support. *Mis Quarterly Executive*, 16(3), 153-169.
- Wilson, H. J., & Daugherty, P. R. (2018). Collaborative intelligence: humans and AI are joining forces. *Harvard Business Review*, 96(4), 114.

- Yeomans, M., Shah, A., Mullainathan, S., & Kleinberg, J. (2019). Making sense of recommendations. *Journal of Behavioral Decision Making*, 32(4), 403-414. doi: 10.1002/bdm.2118
- Zhang, L. X., Pentina, I., & Fan, Y. H. Who do you choose? Comparing perceptions of human vs robo-advisor in the context of financial services. *Journal of Services Marketing*. doi: 10.1108/jsm-05-2020-0162

1 General Discussion

The overarching goal of this thesis was to explore the effects of increasing autonomy on peoples' responses to technical systems. Based on the assumptions that technical autonomy is of essential importance in the digital age and the digital age is characterized by a close collaboration of humans and machines, one conceptual and two empirical contributions were made through this research. The main results and conclusions drawn from them are briefly outlined below.

1.1 Conceptual Contribution

This work provides a more nuanced understanding of advanced technology by identifying two core aspects. First, it clarifies the concepts of autonomy and automation, which often are not clearly distinguished from one another and instead used interchangeably (Vagia, Transeth, & Fjerdings, 2016). Conceptualizing automation and autonomy as separate categories makes it possible to understand advanced technologies in greater depth. Second, the taxonomy developed in this work illustrates that not only technical agency (Rammert, 2008, 2011) but also technical autonomy should be understood as graduated. Based on this notion, the taxonomy establishes different levels of technical autonomy and thus acknowledges and reflects current technical developments. This is useful for representing current technical developments that allow for systems to operate in various ways (Araujo, Helberger, Kruikemeier, & de Vreese, 2020), such as on high and low levels of autonomy. Moreover, the establishment of different levels of autonomy makes possible the operationalization of this technical property, for example by facilitating comparisons of technical systems based on the extent of their autonomy.

The delineation and determination of gradients of automation and autonomy are useful for guiding a number of practical considerations, as demonstrated by Simmler, Weder, and Frischknecht (2021) who applied the taxonomy to legal

procedures. Thus, the taxonomy may assist practitioners in better understanding and anticipating technical systems and their implications for organizational processes. In sum, this taxonomy shows that technical autonomy can be deconstructed into different levels that are important to consider, at least from a normative point of view.

1.2 Empirical Contributions

The taxonomy showed that technical autonomy can indeed be decomposed into different levels from a normative, theoretical perspective. In the introduction and throughout this thesis, the assertion has been made that technical autonomy should be understood from the perspective of those who use and collaborate with such systems. However, whether and how exactly these different levels also affect human responses remains to be addressed.

The second article provides initial support for the notion that technical autonomy creates a shift in social perception, not only from a theoretical perspective but also from a human standpoint. In this research, social perception was operationalized with the ascription of warmth and competence, together comprising the universal dimensions of social cognition (Fiske, Cuddy, & Glick, 2007), as well as with the ascription of experience and agency, together comprising the dimension of mind perception (Gray, Gray, & Wegner, 2007). According to the data, technical autonomy had a statistically significant effect on ascriptions of agency and competence but not warmth and experience. Despite the non-significant findings for warmth and agency, these results suggest that technical autonomy is a variable relevant in the field of human-machine interaction (and collaboration); in very broad and fundamental ways, it affects how a technical system is perceived. Particularly, people ascribed more competence and agency to technical systems operating on higher levels of autonomy (i.e., described as adaptable and/or open), as compared to systems operating on lower levels.

This means that gains in technical autonomy are recognized by humans and the social perception of a particular technical system varies across different levels of autonomy. This provides initial support for the relevance of technical autonomy in the digital age and suggests that technical functioning should be taken into account instead of being regarded as fixed (Sundar, 2020).

Having demonstrated that different levels of technical autonomy may lead to different ascriptions, it is important to also explore more specific responses to this new property, which was the purpose of the third article in this thesis. Based on the evidence presented, this article examined the influence of technical autonomy in organizational decision-making and more specifically on the willingness to accept advice from and delegate decisions to a technical system. The results from the experiment suggest that technical autonomy had a statistically significant effect on both the willingness to accept advice and delegate decisions. More precisely, people were more willing to accept advice from and delegate decisions to a system operating on a high level of autonomy (i.e., described as adaptable and/or open), as compared to systems operating on lower levels.

It is important to point out that these results were not expected. People prefer comprehensible technical systems (Cadario, Longoni & Morewedge, 2021; Yeomans, Shah, Mullainathan & Kleinberg, 2019) while autonomy renders machines opaque and intractable (Rahwan et al., 2019). Thus, people seem to rely on technical systems high in autonomy *although* they are opaque and their outputs difficult to comprehend. The findings indicate that people are more willing to rely on technical systems high in autonomy and provides further support for the relevance of this property. This is a relevant finding for many different areas of technical application and for human-machine collaboration in general.

1.3 Overall Contribution

The findings of both experiments agree with one another and consistently found evidence for the relevance of technical autonomy. This is especially true because increasing levels of autonomy change human responses to technical systems. This change follows a persistent pattern. Autonomous technical systems operating on high levels of autonomy are perceived as having more agency, and more and more people are willing to accept advice and delegate decisions to them. This shift in response manifests most clearly where technical systems are able to learn and are described as open. Current research does not usually account for technical functioning, and rather tends to compare responses to algorithmic decisions to those elicited by human decisions (e.g., Hoddinghaus, Sondern, & Hertel, 2021; Lee, 2018; Yeomans et al., 2019). Moreover, many theories treat technology as a constant and do not explicitly account for technical functioning (Schuetz & Venkatesh, 2020; Sundar, 2020). In conclusion, this thesis provides initial evidence that *the level of technical autonomy is relevant in the digital age* and requires careful consideration in theory and practice.

2 Practical Implications

An ever-increasing number of organizations are now deploying algorithms, algorithmic decision-making systems, and AI (Benbya, Pachidi, & Jarvenpaa, 2021; Jarrahi, 2018; Raisch & Krakowski, 2021; Wilson & Daugherty, 2018). This increased use offers enormous potential, such as in fostering innovation (Brynjolfsson & McAfee, 2012; Raisch & Krakowski, 2021), augmenting service provision (Benbya et al., 2021), and generally improving decision-making (Jarrahi, 2018). The realization of this potential depends on many factors, and a key one is the people who work with these machines. It is the close cooperation between humans and machines that promises the greatest benefits (Jarrahi, 2018; Makarius, Mukherjee, Fox, & Fox, 2020). Therefore, this thesis used decision scenarios to explore how humans react to increasing technical autonomy and such technology being embedded in an organizational context.

The findings show that the level of technical autonomy plays a role in decision-making, as the willingness to accept advice from and delegate decisions to a technical system was stronger when that system was described as adaptable or both adaptable and open.

If companies want to promote the use of highly autonomous technical systems, it would be useful to highlight the advanced capabilities of these systems. In this research, the adaptability (i.e., machine learning) and openness of the system most increased the willingness to depend on such a system. Moreover, while technical autonomy was determined to alter subjects' perceptions of competence and agency, it had had no effect on warmth and experience. The perception of warmth was especially closely tied to trust (Cuddy, Glick, & Beninger, 2011; Glikson & Woolley, 2020; Kulms & Kopp, 2018). Thus, to promote trust in such systems, it would be helpful to focus not only on technical capabilities (such as machine learning and openness) but also on factors that promote the perceived warmth of a system.

It is not, however, beneficial for companies if employees and managers rely solely or too heavily on these systems when making decisions. In many cases, it is the combination of human intuition and the machine's immense data processing ability that offers the greatest likelihood of success (Jarrahi, 2018). For companies, this means that users need to be well-trained and thoroughly educated with regards to the benefits of such machines, what they should be used for, and how exactly they should be employed. Thus, public and private sector organizations should establish clear guidelines and educate their employees on these topics.

Because the findings indicate that people quite readily leave certain decisions entirely to autonomous technical systems, it is important to carefully determine what tasks are suitable for such machines. Autonomous technical systems will not be accepted equally for use with all tasks, especially when such tasks require human judgement and intuition (Castelo, Bos, & Lehmann, 2019; Lee, 2018). For example, Lee (2018) found that an algorithm evaluating work performance was

regarded as unfair and untrustworthy. From a strategic point of view, it is also important to thoroughly evaluate what tasks can be left entirely or partly to increasingly autonomous technical systems. Technical autonomy makes systems opaque and their outputs difficult to understand (Rahwan et al., 2019). Decisions made by systems with substantial autonomy are therefore difficult to explain and comprehend, and raise questions related to accountability (Zarsky, 2016). Especially when decisions must be justified to shareholders or the public, it is worthwhile not to use autonomous machines and rather leave decisions to humans (i.e., responsible managers) alone.

Technical systems with a high degree of autonomy are attributed a greater ability to act than are technical systems operating at lower levels of autonomy. The perception of agency is closely linked to the attribution of responsibility (Gray et al., 2007; Waytz, Gray, Epley, & Wegner, 2010). In an organizational context, this could mean that people no longer consider themselves responsible for (negative) outcomes, and rather delegate responsibility to autonomous technical systems. Once again, it will be important for organizations to establish clear guidelines regarding how to use such systems and clarify who is accountable for what.

3 Limitations

Like every research endeavor, this project has its limitations. One concern is related to the research approach. The data from which the conclusions were drawn were gathered via online experiments, and thus were not collected in a realistic setting. Due to the strict control of possible confounding factors, experiments generally have high internal validity (Bortz & Döring, 2006), meaning they are indicative of the highly specific situations created in the experiment. However, this is at the expense of the external validity and generalizability of the results to situations outside the experiment in particular and laboratory conditions in general (Bortz & Döring, 2006). Hence, the findings obtained from the experiments conducted for this research should be

interpreted with caution. Whether and how the findings are generalizable to and apply in real-world situations remains to be seen. Furthermore, in both experiments, technical autonomy (and thus only one specific variable) was under investigation. In more realistic settings, however, human-machine collaboration takes place in highly complex environments where technical autonomy is likely to be only one of many influential factors. Hence, the actual effects of technical autonomy on social perceptions and system reliance are likely to be overestimated. Moreover, the responses to autonomous technologies were not assessed with behavioral measures, meaning that this thesis could not draw conclusions about how people will behave when interacting or collaborating with increasingly autonomous machines. Likewise, no participants had interacted or collaborated with an actual technical system, neither high nor low in autonomy. The experimental manipulation of technical autonomy was based wholly on scenarios and thus subject to respondents' imagination.

The lack of a specific context is another limitation, as human responses were explored in either the complete absence of contextual information or generic organizational decision-making situations. However, it can be assumed that technical systems have to meet different requirements and expectations if they are used in the public rather than private sector. In the public sector, the accountability requirements are quite different (Diakopoulos, 2016). Likewise, in medicine, there are likely to be different requirements and expectations than in retail. The extent to which the results can be generalized to other contexts is therefore questionable. However, without deriving and testing the basic premises that technical autonomy is important in the digital age and increased technical autonomy alters the human response to technical systems, no applied or more practice-oriented research could take place.

This project investigated technical autonomy from the perspective of humans interacting and collaborating with increasingly autonomous technical systems. Yet, people are also the recipients of algorithmic outputs (Nagtegaal, 2021) (i.e., when not involved in the process of making a decision). The question of how

people such as citizens, customers, and patients respond to definite decisions made by increasingly autonomous machines remains unanswered.

Finally, this thesis shows that technical autonomy influences certain human responses. However, it cannot resolve why differences in responses appear. Deductive, hypothesis-driven research is only able to explain phenomena of such complexity to a limited extent.

4 Areas for Further Research

This project is basic research focused on exploring whether technical autonomy has an influence on human reception. Thus, the main goal was to test a basic premise. This means that (as implied in the limitations), there is tremendous potential for subsequent applied research. Fruitful areas for future work are outlined below.

First, the data obtained, and findings presented here are not rooted in a specific context. However, considering the context will be important in future research because human responses and ascriptions can differ based on the task a machine fulfills (see, e.g., Castelo et al., 2019; Lee, 2018; Wang & Krumhuber, 2018). At the same time, the context itself may affect how such machines are used. For example, one could assume that increasingly autonomous machines are handled with more caution in the public rather than private sector because the former requires more accountability (and thus explanations) than the latter (Diakopoulos, 2016). For the same reason, one could also assume that public managers are more cautious than private sector managers because they must explain their decisions to the public. It would be useful to examine the role of technical autonomy on a more collective level. In many cases, and especially in business organizations, decisions are the result of collective processes and not made by individuals (Glikson & Woolley, 2020). Investigating how groups of people make sense of and incorporate advice from increasingly autonomous machines would be a promising field of future research.

Second, this research investigated the influence of only one variable (i.e., technical autonomy). However, there are several additional variables that function in human-machine collaboration. Future research could, for example, explore whether there are age effects in the social perception of and response to increasingly autonomous technologies. Age effects have been found to be relevant to the adoption of technical systems (Venkatesh, Thong, & Xu, 2012). Another important factor to consider is time. Trust in technical systems declines over time; however, it is not clear if increasing levels of technical autonomy might prevent this decline (Glikson & Woolley, 2020). Likewise, the design of a specific program or its interface could be as important as its actual technical capabilities in terms of establishing trust in the user. In sum, further research should add a broad set of variables to the experimental design and further investigate how these variables interact with technical autonomy.

Lastly, this thesis was written from a positivist view, meaning that it only describes and suggests how people may perceive increasingly autonomous technologies. Complementing this view with a more normative approach would be another potentially fruitful area for further research. This kind of research could, for example, determine situations in which high levels of autonomy are appropriate and where low levels of autonomy should instead be prioritized.

5 References

- Araujo, T., Helberger, N., Kruikemeier, S., & de Vreese, C. H. (2020). In AI we trust? Perceptions about automated decision-making by artificial intelligence. *Ai & Society*, 35(3), 611-623. doi: 10.1007/s00146-019-00931-w
- Benbya, H., Pachidi, S., & Jarvenpaa, S. L. (2021). Special Issue Editorial: Artificial Intelligence in Organizations: Implications for Information Systems Research. *Journal of the Association for Information Systems*, 22(2), 281-303. doi: 10.17705/1jais.00662
- Bortz, J., & Döring, N. (2006). Empirische Forschung im Überblick. In J. Bortz & N. Döring (Eds.), *Forschungsmethoden und Evaluation: für Human- und Sozialwissenschaftler* (pp. 1-34). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Brynjolfsson, E., & McAfee, A. (2012). Winning the Race With Ever-Smarter Machines. *Mit Sloan Management Review*, 53(2), 53-60.
- Cadario, R., Longoni, C., & Morewedge, C.K. (2021). Understanding, explaining, and utilizing medical artificial intelligence. *Nature Human Behaviour*.
- Castelo, N., Bos, M. W., & Lehmann, D. R. (2019). Task-Dependent Algorithm Aversion. *Journal of Marketing Research*, 56(5), 809-825. doi: 10.1177/0022243719851788
- Cuddy, A. J. C., Glick, P., & Beninger, A. (2011). The dynamics of warmth and competence judgments, and their outcomes in organizations. *Research in Organizational Behavior*, 31, 73-98. doi: <https://doi.org/10.1016/j.riob.2011.10.004>
- Diakopoulos, N. (2016). Accountability in Algorithmic Decision Making. *Communications of the Acm*, 59(2), 56-62. doi: 10.1145/2844110
- 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. doi: 10.1016/j.tics.2006.11.005
- Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical evidence. *Academy of Management Annals*, 14(2), 627-660. doi: 10.5465/annals.2018.0057
- Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science*, 315(5812), 619-619. doi: 10.1126/science.1134475
- Hoddinghaus, M., Sondern, D., & Hertel, G. (2021). The automation of leadership functions: Would people trust decision algorithms? *Computers in Human Behavior*, 116. doi: 10.1016/j.chb.2020.106635

-
- Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business Horizons*, 61(4), 577-586. doi: 10.1016/j.bushor.2018.03.007
- Kulms, P., & Kopp, S. (2018). A Social Cognition Perspective on Human-Computer Trust: The Effect of Perceived Warmth and Competence on Trust in Decision-Making With Computers. *Frontiers in Digital Humanities*, 5(14). doi: 10.3389/fdigh.2018.00014
- Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*, 5(1), 2053951718756684. doi: 10.1177/2053951718756684
- Makarius, E. E., Mukherjee, D., Fox, J. D., & Fox, A. K. (2020). Rising with the machines: A sociotechnical framework for bringing artificial intelligence into the organization. *Journal of Business Research*, 120, 262-273. doi: 10.1016/j.jbusres.2020.07.045
- Nagtegaal, R. (2021). The impact of using algorithms for managerial decisions on public employees' procedural justice. *Government Information Quarterly*, 38(1). doi: 10.1016/j.giq.2020.101536
- Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J.-F., Breazeal, C., . . . Wellman, M. (2019). Machine behaviour. *Nature*, 568(7753), 477-486. doi: 10.1038/s41586-019-1138-y
- Raisch, S., & Krakowski, S. (2021). Artificial Intelligence and Management: The Automation-Augmentation Paradox. *The Academy of Management Review*, 46(1), 192. doi: 10.5465/amr.2018.0072
- Rammert, W. (2008). Where the action is: Distributed agency between humans, machines and programs. Berlin.
- Rammert, W. (2011) Distributed Agency and Advanced Technology Or: How to Analyse Constellations of Collective Inter-Agency. *Technical University Technology Studies Working Papers*. Berlin.
- Schuetz, S., & Venkatesh, V. (2020). Research Perspectives: The Rise of Human Machines: How Cognitive Computing Systems Challenge Assumptions of User-System Interaction. *Journal of the Association for Information Systems*, 21(2), 460-482. doi: 10.17705/1jais.00608
- Simmler, M., Weder, R., & Frischknecht, R. (2021). Rechtliche Verfahren als sozio-technische Systeme: Eine Taxonomie des Zusammenwirkens von Mensch und Maschine in der Rechtsanwendung. In M. Simmler (Ed.), *Smart Criminal Justice. Der Einsatz von Algorithmen in der Polizeiarbeit und Strafrechtspflege* (pp. 81-99). Basel: Helbing Lichtenhahn Verlag.
- Sundar, S. S. (2020). Rise of Machine Agency: A Framework for Studying the Psychology of Human-AI Interaction (HAI). *Journal of Computer-Mediated Communication*, 25(1), 74-88. doi: 10.1093/jcmc/zmz026

- Vagia, M., Transeth, A. A., & Fjerdings, S. A. (2016). A literature review on the levels of automation during the years. What are the different taxonomies that have been proposed? *Applied Ergonomics*, 53, 190-202. doi: 10.1016/j.apergo.2015.09.013
- Venkatesh, V., Thong, J. Y. L., & Xu, X. (2012). Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. *MIS Quarterly*, 36(1), 157-178.
- Wang, X. J., & Krumhuber, E. G. (2018). Mind Perception of Robots Varies With Their Economic Versus Social Function. *Frontiers in Psychology*, 9. doi: 10.3389/fpsyg.2018.01230
- Waytz, A., Gray, K., Epley, N., & Wegner, D. M. (2010). Causes and consequences of mind perception. *Trends in Cognitive Sciences*, 14(8), 383-388. doi: 10.1016/j.tics.2010.05.006
- Wilson, H. J., & Daugherty, P. R. (2018). Collaborative intelligence: humans and AI are joining forces. *Harvard Business Review*, 96(4), 114.
- Yeomans, M., Shah, A., Mullainathan, S., & Kleinberg, J. (2019). Making sense of recommendations. *Journal of Behavioral Decision Making*, 32(4), 403-414. doi: 10.1002/bdm.2118
- Zarsky, T. (2016). The Trouble with Algorithmic Decisions: An Analytic Road Map to Examine Efficiency and Fairness in Automated and Opaque Decision Making. *Science Technology & Human Values*, 41(1), 118-132. doi: 10.1177/0162243915605575

Appendix B: A note about the linearity of the levels of technical autonomy

The results sections of the experiments presented before do not provide in-depth analyses of how the theoretically derived and experimentally manipulated levels of technical autonomy correspond with perceived autonomy (i.e., the amount of autonomy the participants ascribed to the system). Thus, the following additional data are included to demonstrate that the manipulations in both experiments functioned appropriately and indicated that with an increasing level of autonomy, the technical system tended to be perceived as more autonomous.

In the first experiment, with the level of autonomy as a factor and perceived autonomy as the dependent variable, a robust one-way ANOVA applying Welch's procedure showed a statistically significant main effect ($F(4,116.4) = 3.21, p = .015, \eta_p^2 = .047$). There was a statistically significant linear trend ($t(134) = 2.77, p = .006, d = 1.288$), indicating that the more autonomous a technical system was described as being, the more autonomous it was perceived to be by the participants. A similar pattern was observed in the second experiment. A robust Welch ANOVA was run, with the level of technical autonomy as a factor and perceived amount of autonomy as the dependent variable. This test revealed a statistically significant effect of the manipulation ($F(4,140) = 3.15, p = .016, \eta_p^2 = .041$). Again, there was a statistically significant linear trend detected ($t(167) = 2.39, p = .018, d = .945$).

This linear trend was further highlighted by a correlational analysis of the two samples. In the first experiment, there was a statistically significant relationship (obtained by applying Spearman's rho) between the manipulated level of technical autonomy and perceived autonomy ($r_s = 0.175, p = .007$). However, after conducting the same analysis in the second experiment, only a marginally statistically significant relationship between the level of technical autonomy and perceived autonomy was found ($r_s = .113, p = .058$).

Although the data regarding the relationship between the theoretically assumed and perceived levels of autonomy were not perfectly exemplary, they still provided evidence that the manipulation worked. Together, these analyses suggest that the higher the level of (theoretically assumed and experimentally manipulated) technical autonomy, the more likely these systems are to be perceived as autonomous.

Curriculum Vitae

Ruth Frischknecht

* June 6, 1990 in Herisau

Education

2017 – 2022	PhD in Management , Specialized in Business Innovation, University of St.Gallen
2013 – 2015	M.Sc. in Psychology , Major in Social Psychology, Economic Psychology and Decision Science, University of Basel
2009 – 2013	B.Sc. in Psychology , University of Basel
2003 – 2009	Matura , Gymnasium St. Antonius, Appenzell

Work Experience

From 9/2022	Project Leader , Political Planning, Controlling and Management Support, State Chancellery, Canton of St.Gallen
10/2016 – 7/2022	Researcher and Teaching Assistant , Institute for Systemic Management and Public Governance, University of St.Gallen
2/2015 – 10/2016	Traineeship , St.Galler Tagblatt