

The Effects of User Assistance Systems on User Perception and Behavior

Dissertation

zur Erlangung des akademischen Grades

Dr.-Ing.

eingereicht an der

Mathematisch-Naturwissenschaftlich-Technischen Fakultät

der Universität Augsburg

von

Christian Olenberger, M.Sc. with honors

Augsburg, März 2023



Erstgutachter:

Prof. Dr. Henner Gimpel

Zweitgutachter:

Prof. Dr. Elisabeth André

Datum der Disputation:

26. Mai 2023

*“If we want users to like our software, we should design it to behave like a likeable person:
respectful, generous, and helpful.”*

Alan Cooper

Abstract

The rapid development of information technology (IT) is changing how people approach and interact with IT systems (Maedche et al. 2016). IT systems can increasingly support people in performing ever more complex tasks (Vtyurina and Fourney 2018). However, people's cognitive abilities have not evolved as quickly as technology (Maedche et al. 2016). Thus, different external factors (e.g., complexity or uncertainty) and internal conditions (e.g., cognitive load or stress) reduce decision quality (Acciarini et al. 2021; Caputo 2013; Hilbert 2012). User-assistance systems (UASs) can help to compensate for human weaknesses and cope with new challenges. UASs aim to improve the user's cognition and capabilities, benefiting individuals, organizations, and society. To achieve this goal, UASs collect, prepare, aggregate, analyze information, and communicate results according to user preferences (Maedche et al. 2019). This support can relieve users and improve the quality of decision-making.

Using UASs offers many benefits but requires successful interaction between the user and the UAS. However, this interaction introduces social and technical challenges, such as loss of control or reduced explainability, which can affect user trust and willingness to use the UAS (Maedche et al. 2019). To realize the benefits, UASs must be developed based on an understanding and incorporation of users' needs. Users and UASs are part of a socio-technical system to complete a specific task (Maedche et al. 2019). To create a benefit from the interaction, it is necessary to understand the interaction within the socio-technical system, i.e., the interaction between the user, UAS, and task, and to align the different components.

For this reason, this dissertation aims to extend the existing knowledge on UAS design by better understanding the effects and mechanisms during the interaction between UASs and users in different application contexts. Therefore, theory and findings from different disciplines are combined and new theoretical knowledge is derived. In addition, data is collected and analyzed to validate the new theoretical knowledge empirically. The findings can be used to reduce adaptation barriers and realize a positive outcome. Overall this dissertation addresses the four classes of UASs presented by Maedche et al. (2016): basic UASs, interactive UASs, intelligent UASs, and anticipating UASs.

First, this dissertation contributes to understanding how users interact with *basic UASs*. Basic UASs do not process contextual information and interact little with the user (Maedche et al. 2016). This behavior makes basic UASs suitable for application contexts, such as social media, where little interaction is desired. Social media is primarily used for entertainment and focuses

on content consumption (Moravec et al. 2018). As a result, social media has become an essential source of news but also a target for fake news, with negative consequences for individuals and society (Clarke et al. 2021; Laato et al. 2020). Thus, this thesis presents two approaches to how basic UASs can be used to reduce the negative influence of fake news. Firstly, basic UASs can provide interventions by warning users of questionable content and providing verified information but the order in which the intervention elements are displayed influences the fake news perception. The intervention elements should be displayed after the fake news story to achieve an efficient intervention. Secondly, basic UASs can provide social norms to motivate users to report fake news and thereby stop the spread of fake news. However, social norms should be used carefully, as they can backfire and reduce the willingness to report fake news.

Second, this dissertation contributes to understanding how users interact with *interactive UASs*. Interactive UASs incorporate limited information from the application context but focus on close interaction with the user to achieve a specific goal or behavior (Maedche et al. 2016). Typical goals include more physical activity, a healthier diet, and less tobacco and alcohol consumption to prevent disease and premature death (World Health Organization 2020). To increase goal achievement, previous researchers often utilize digital human representations (DHRs) such as avatars and embodied agents to form a socio-technical relationship between the user and the interactive UAS (Kim and Sundar 2012a; Pfeuffer et al. 2019). However, understanding how the design features of an interactive UAS affect the interaction with the user is crucial, as each design feature has a distinct impact on the user's perception. Based on existing knowledge, this thesis highlights the most widely used design features and analyzes their effects on behavior. The findings reveal important implications for future interactive UAS design.

Third, this dissertation contributes to understanding how users interact with *intelligent UASs*. Intelligent UASs prioritize processing user and contextual information to adapt to the user's needs rather than focusing on an intensive interaction with the user (Maedche et al. 2016). Thus, intelligent UASs with emotional intelligence can provide people with task-oriented and emotional support, making them ideal for situations where interpersonal relationships are neglected, such as crowd working. Crowd workers frequently work independently without any significant interactions with other people (Jäger et al. 2019). In crowd work environments, traditional leader-employee relationships are usually not established, which can have a negative impact on employee motivation and performance (Cavazotte et al. 2012). Thus, this thesis examines the impact of an intelligent UAS with leadership and emotional capabilities on employee performance and enjoyment. The leadership capabilities of the intelligent UAS lead

to an increase in enjoyment but a decrease in performance. The emotional capabilities of the intelligent UAS reduce the stimulating effect of leadership characteristics.

Fourth, this dissertation contributes to understanding how users interact with *anticipating UASs*. Anticipating UASs are intelligent and interactive, providing users with task-related and emotional stimuli (Maedche et al. 2016). They also have advanced communication interfaces and can adapt to current situations and predict future events (Knote et al. 2018). Because of these advanced capabilities anticipating UASs enable collaborative work settings and often use anthropomorphic design cues to make the interaction more intuitive and comfortable (André et al. 2019). However, these anthropomorphic design cues can also raise expectations too high, leading to disappointment and rejection if they are not met (Bartneck et al. 2009; Mori 1970). To create a successful collaborative relationship between anticipating UASs and users, it is important to understand the impact of anthropomorphic design cues on the interaction and decision-making processes. This dissertation presents a theoretical model that explains the interaction between anthropomorphic anticipating UASs and users and an experimental procedure for empirical evaluation. The experiment design lays the groundwork for empirically testing the theoretical model in future research.

To sum up, this dissertation contributes to information systems knowledge by improving understanding of the interaction between UASs and users in different application contexts. It develops new theoretical knowledge based on previous research and empirically evaluates user behavior to explain and predict it. In addition, this dissertation generates new knowledge by prototypically developing UASs and provides new insights for different classes of UASs. These insights can be used by researchers and practitioners to design more user-centric UASs and realize their potential benefits.

Table of Contents

Abstract	II
Table of Contents	V
List of Figures	VIII
List of Tables	X
List of Abbreviations	XII
1. Introduction	1
1.1. Motivation	1
1.2. The Role of User Assistance Systems in Information Systems Research	3
1.3. Aim and Outline of this Dissertation	7
2. Theoretical and Methodological Background	14
2.1. User Assistance Systems	14
2.2. Research Methods in Information Systems Research	19
3. Basic User Assistance Systems to Combat Fake News	23
3.1. Theoretical Background	24
3.2. The Effect of Intervention Order on Fake News Perception	28
3.3. The Effect of Social Norms on Fake News Reporting	49
4. Interactive User Assistance Systems to Promote Health Behavior Change	74
4.1. Theoretical Background	75
4.2. Methodology	79
4.3. Results	81
4.4. Discussion	101
4.5. Conclusion	109
5. Intelligent User Assistance Systems to Enhance Crowd Working	110
5.1. Theoretical Background	113
5.2. Hypotheses Development	116
5.3. Methodology	121
5.4. Results	129
5.5. Discussion	132
5.6. Conclusion	137
6. Anticipating User Assistance Systems to Improve Collective Decision-Making	139

6.1.	Theoretical Background	141
6.2.	Hypothesis Development.....	143
6.3.	Potential Experiment Design.....	147
7.	General Discussion and Conclusion.....	152
7.1.	Summary of Results and Meta-Inferences.....	152
7.2.	Conclusion.....	164
	References	166
Appendix A	Basic User Assistance Systems to Combat Fake News.....	212
Appendix A.1	Detailed Sample Description	212
Appendix A.2	Results Regarding Structural Equality of Treatments.....	214
Appendix A.3	Additional Control Treatments.....	215
Appendix A.4	Stimuli	216
Appendix A.5	List of Scales	220
Appendix A.6	Descriptive Overview With All Treatments.....	224
Appendix A.7	Descriptive Overview of the Data	226
Appendix A.8	Detailed Report of Results.....	228
Appendix A.9	Effects of Warning Labels Extended by Verified Information.....	233
Appendix A.10	Detailed Sample Description	235
Appendix A.11	Stimuli	237
Appendix A.12	Results Regarding Structural Equality of Treatments.....	243
Appendix A.13	Development and Validation of the Newsfeed	244
Appendix A.14	List of Scales	249
Appendix A.15	Results of the Brant test for Ordered Logistic Regression.....	254
Appendix A.16	Results of Real News Reporting.....	255
Appendix A.17	Results of the Ordered Logistic Regression Including Awareness.	256
Appendix A.18	Effects of Strength of Descriptive Social Norms	258
Appendix B	Interactive User Assistance Systems to Promote Behavior Change	259
Appendix B.1	Overview of all Reviewed Papers.....	259
Appendix B.2	Summary of Risk of Bias Analysis.....	265
Appendix B.3	Detailed Results of Risk of Bias Analysis.....	268
Appendix C	Intelligent User Assistance System to Improve Working Quality	271
Appendix C.1	Leadership Capabilities of Modern Conversational Agents	271
Appendix C.2	Interaction Screens With Conversational Agent.....	274

Appendix C.3	Experimental Treatment Characteristics and Capabilities	275
Appendix C.4	Exemplary Treatment Capabilities	276
Appendix C.5	Measurement Items.....	277
Appendix C.6	Measurement Characteristics	278
Appendix C.7	Test for Common-Method Variance	280
Appendix C.8	Manipulation Check.....	281

List of Figures

<i>Figure 1: Socio-Technical System (Based on Maedche et al. (2019))</i>	4
<i>Figure 2: User Assistance Systems Classification (Based on Maedche et al. (2016))</i>	6
<i>Figure 3: Structure of This Dissertation</i>	13
<i>Figure 4: Simplified Overview of the Study Procedure</i>	35
<i>Figure 5: Overview of the Intervention Design in Each Treatment</i>	37
<i>Figure 6: Summary of the Procedure of the Study</i>	40
<i>Figure 7: Comparison of the Mean Believability of Fake News</i>	41
<i>Figure 8: Mental Model Measurement per Treatment</i>	42
<i>Figure 9: Implementation of the Injunctive and Descriptive SN Messages</i>	57
<i>Figure 10: Summary of the Procedure of the Experiment in Study 1</i>	59
<i>Figure 11: BCIs and COM-B Model (Adapted from Michie et al. (2011))</i>	77
<i>Figure 12: Structure Underlying the Present Review</i>	79
<i>Figure 13: Stages of the Structured Literature Review</i>	79
<i>Figure 14: Overview of the Results</i>	100
<i>Figure 15: Research Model</i>	117
<i>Figure 16: Summary of the Experimental Procedure</i>	122
<i>Figure 17: Summary of the Three Experimental Treatment Phases</i>	124
<i>Figure 18: Representation of the Female and Male Conversational Agent</i>	125
<i>Figure 19: Structural Equation Model of the Research Model</i>	130
<i>Figure 20: Research Model</i>	147
<i>Figure 21: Sequence of Steps to Be Performed by Each Participant</i>	151
<i>Figure 22: Post Design in Additional Control Treatments</i>	215
<i>Figure 23: Fake News Stimuli</i>	217
<i>Figure 24: Real News Stimuli During Fake News Exposure</i>	218
<i>Figure 25: Real News Stimuli During Article Recognition</i>	219
<i>Figure 26: Comparison of the Mean Believability of Fake News</i>	224
<i>Figure 27: Mental Model Measurement per Treatment</i>	225
<i>Figure 28: Fake News Posts in Study 1</i>	237
<i>Figure 29: Real News Posts in Study 1</i>	238
<i>Figure 30: Neutral Posts in Study 1</i>	239
<i>Figure 31: Fake News Posts in Study 2</i>	240
<i>Figure 32: Real News Posts in Study 2</i>	241

Figure 33: Neutral Posts in Study 2..... 242
Figure 34: Interaction Screens in Two Experimental Treatment Phases 274
Figure 35: Exemplary Treatment Capabilities 276

List of Tables

<i>Table 1: Results of the Believability Comparison</i>	43
<i>Table 2: Results of the Adaptation of Fake News</i>	44
<i>Table 3: Results of the Adaptation of Verified Information</i>	45
<i>Table 4: Overview of Hypotheses and Empirical Results</i>	45
<i>Table 5: Comparison of Fake News Reporting in Study 1</i>	61
<i>Table 6: Probability of News Post Reporting Based on the Descriptive SN</i>	64
<i>Table 7: Comparison of Fake News Reporting in Study 2</i>	66
<i>Table 8: Motivation Factors and Hurdles to Report Fake News</i>	67
<i>Table 9: Psychological Constructs Evaluated in the 60 Papers in the Review</i>	89
<i>Table 10: Intervention Functions and Relationship to Psychological Constructs</i>	95
<i>Table 11: Targeted SNAP Health Behaviors, Methods, and Measures</i>	98
<i>Table 12: Mean and Standard Deviation in the Experimental Treatment Conditions</i>	129
<i>Table 13: Summary of Hypotheses Testing</i>	131
<i>Table 14: Definition of Core Constructs</i>	142
<i>Table 15: Exemplary Configuration of Anthropomorphic Cues</i>	150
<i>Table 16: Detailed Sample Description of Phase 1</i>	212
<i>Table 17: Detailed Sample Description of Phase 2</i>	213
<i>Table 18: Structural Equality of the Characteristics of the Participants in Phase 1</i>	214
<i>Table 19: Structural Equality of the Characteristics of the Participants in Phase 2</i>	214
<i>Table 20: Description of All Scales Used</i>	223
<i>Table 21: Comparison of Mean Believability Measurement for Fake News Posts</i>	226
<i>Table 22: Comparison of Number of Mental Models Based on Fake News Posts</i>	226
<i>Table 23: Comparison of Number of Mental Models Based on Verified Information</i>	227
<i>Table 24: Results of Believability Comparison</i>	228
<i>Table 25: Results of Fake News Adoption</i>	229
<i>Table 26: Results of Verified Information Adoption</i>	230
<i>Table 27: Results of Memory of the Fake News Articles</i>	231
<i>Table 28: Results of Memory of the Warning Label</i>	232
<i>Table 29: Results of Memory of the Verified Information</i>	232
<i>Table 30: Detailed Sample Description of Study 1</i>	235
<i>Table 31: Detailed Sample Description of Study 2</i>	236
<i>Table 32: Structural Equality of the Characteristics of the Participants in Study 1</i>	243

<i>Table 33: Structural Equality of the Characteristics of the Participants in Study 2</i>	243
<i>Table 34: Number of Reports per Post in the Fake Assessment Run of Study 1</i>	246
<i>Table 35: Overall Number of Reports in the Fake Assessment Run of Study 1</i>	246
<i>Table 36: Summary of Control Items in Study 2</i>	247
<i>Table 37: Number of News Stories Rated as Fake News in Study 2.</i>	248
<i>Table 38: Overall Number of News Stories Rated as Fake News in Study 2</i>	248
<i>Table 39: Description of all Scales Used in Study 1</i>	250
<i>Table 40: Description of All Scales Used in Study 2</i>	253
<i>Table 41: Results of the Brant Test in Study 1</i>	254
<i>Table 42: Results of the Brant Test in Study 2</i>	254
<i>Table 43: Results of Real News Reporting of Study 1</i>	255
<i>Table 44: Results of Fake News Reporting While Controlling For Awareness</i>	256
<i>Table 45: Results of Fake News Reporting While Controlling for Awareness</i>	257
<i>Table 46: Effects of Strength of Descriptive Social Norms on Fake News in Study 1</i>	258
<i>Table 47: Effects of Strength of Descriptive Social Norms on Real News in Study 1</i>	258
<i>Table 48: Summary of all Reviewed Papers</i>	264
<i>Table 49: Summary of Risk of Bias Analysis Across the Five Different RoB2 Areas</i>	266
<i>Table 50: Detailed Results of Risk of Bias Analysis with Risk of Bias 2 Tool</i>	270
<i>Table 51: Exemplary Conversation Between a CA (ChatGPT) and a Crowd worker</i>	273
<i>Table 52: Experimental Treatment Characteristics and Capabilities</i>	275
<i>Table 53: Measurement Items</i>	277
<i>Table 54: Outer Loadings and Cross-Loadings</i>	278
<i>Table 55: Validity and Reliability Characteristics of the Measurement Model</i>	279
<i>Table 56: Results of Marker Variable Analysis for Common-Method Variance</i>	280

List of Abbreviations

AI.....	<i>Artificial Intelligence</i>
BCI.....	<i>Behavior Change Intervention</i>
CA.....	<i>Conversational Agent</i>
COM-B.....	<i>Capability, Opportunity, Motivation, Behavior</i>
DHR.....	<i>Digital Human Representation</i>
EPT.....	<i>Experimental Treatment Phase</i>
HIT.....	<i>Human-Intelligence Task</i>
IS.....	<i>Information Systems</i>
IT.....	<i>Application Programming Interface, Information Technology</i>
LMX.....	<i>Leader-Member Exchange</i>
MTurk.....	<i>Amazon Mechanical Turk</i>
PLS-SEM.....	<i>Partial Least Square Structural Equation Modeling</i>
SEM.....	<i>Structural Equation Modeling</i>
SN.....	<i>Social Norm</i>
SNAP.....	<i>Smoking, Nutrition, Alcohol Overconsumption, Physical Inactivity</i>

1. Introduction

1.1. Motivation

Information technology (IT) capabilities are changing rapidly (Maedche et al. 2016). This development changes how people use, interact and co-exist with IT systems (Mirbabaie et al. 2021; Seeber et al. 2020; Siemon 2022). Traditionally, IT has been used as a tool to accomplish specific predefined tasks (Siemon 2022). However, technological advancements are creating new opportunities for human-machine interaction. The achievements in artificial intelligence (AI) allow IT systems to be used in more sophisticated scenarios (Vtyurina and Fourney 2018; Winkler et al. 2019), whereby IT systems increasingly assist people in performing various tasks. This assistance is more important than ever in today's complex world. Compared to the high pace at which IT systems have advanced, people's cognitive abilities have not developed so quickly (Maedche et al. 2016). There are various external and internal factors, such as complexity, uncertainty, longevity, and perceived irrelevance, but also cognitive load, laziness, stress, and anxiety, which lead individuals to rely on decision-making heuristics (Acciarini et al. 2021; Caputo 2013; Hilbert 2012). These heuristics introduce cognitive biases that lead to systematic errors in decision-making (Gilovich et al. 2002; Weinmann et al. 2016). Today's complex world further reinforces this phenomenon. First, the increasing digitalization of our world and the daily use of IT means that individuals are exposed to more information. This can lead to the individual being exposed to too much information, making them feel overwhelmed. Information overload can result in worse decisions and potential benefits not being realized (Pirkkalainen and Salo 2016). Second, decision-making is becoming more complex as more information is available that needs to be included in the decision-making process. This decision-making process can involve different activities such as information acquisition, validation, filtering, and analysis which further increases the cognitive load placed on individuals. Therefore, IT systems need to assist people in compensating for individual vulnerabilities and coping with the new challenges to make well-founded decisions.

For this purpose, particularly user assistance systems (UASs) have gained attention (Maedche et al. 2016; Pfeuffer et al. 2019). UASs aim to improve the perception, cognition, and capabilities of the user (Jalaliniya and Pederson 2015; Knote et al. 2019) and thereby improve the quality of both private and work-life (Knote et al. 2018). UASs can be integrated into existing IT systems to support users in searching, preparing, processing, or providing

information. For example, researchers introduced different applications to automatically detect false information on social media. These applications assist users during the consumption of social media content by warning them of incorrect information and automatically providing verified information (Gimpel et al. 2020; Kim and Dennis 2019; Moravec et al. 2020). At the same time, UASs are not only becoming more intelligent and capable of solving increasingly complex problems, but they are also becoming more interactive and adopt more and more human-like characteristics, such as empathy, emotions, and personality (Siemon 2022). These capabilities unlock additional application areas and opportunities for UASs to assist humans and reduce the complexity of daily life (Knote et al. 2019). For example, text-based conversational agents (CAs) are increasingly being implemented to improve service availability for users, provide new services, and simplify service offerings through natural language interaction. After the Covid-19 outbreak, the Indian government released a WhatsApp chatbot that could answer users' questions quickly and intuitively to counter uncertainty and misinformation (George et al. 2021a). According to the bot's usage data, 85 million users interacted with the chatbot (Das and Shinde 2022). Modern voice assistants, such as Siri, Alexa, or Google Assistant, interact with the user via natural language to increase the intuitiveness of the communication between the UAS and the user (Knote et al. 2019). In addition to these software innovations, new hardware capabilities fuel the availability of UASs. Smart devices, such as smart home devices, smartphones, or connected cars, enable access to UASs at any time and place, contributing to the growing omnipresence of UASs (Feine et al. 2019; Maedche et al. 2019).

The use of UASs offers a wide variety of benefits. First, individuals benefit from UASs because UASs can search, filter, sort, preprocess, and provide additional information (Maedche et al. 2019). This support can reduce the cognitive load and improve decision quality. At the same time, new forms of interaction, such as voice assistants, enable access to users that may find traditional interfaces challenging due to physical or mental limitations (Seiderer et al. 2020; Vieira et al. 2022). Second, organizations benefit from using UASs because the quality and availability of their services can be improved by automation (Zierau et al. 2020). This results in resource savings (Carter and Knol 2019; Dhiman et al. 2022), increased customer satisfaction (Følstad and Skjuve 2019), and reduced employee workload (Carter and Knol 2019; Meyer von Wolff et al. 2020). Third, society may also benefit from the use of UASs. By improving individual decisions, large-scale benefits can be realized. For example, UASs can help individuals to handle false information better. If enough individuals use these UASs, this will

have a positive impact on society as a whole and can counteract the negative consequences of false information that are currently observable (Clarke et al. 2021; Laato et al. 2020; Roozenbeek and van der Linden 2019; van der Linden et al. 2017).

These benefits can only be realized if people use UASs successfully. However, the interaction between humans and UASs involves various social and technical challenges (Morana et al. 2020). First, UASs take over tasks from users, which can be convenient on the one hand, but on the other hand, it can lead to a sense of loss of control among users (André et al. 2019). Second, the increasing complexity of UASs reduces the explainability and interpretability of the results and actions (Maedche et al. 2019). These phenomena have a negative impact on user trust and willingness to use (Maedche et al. 2019). As a result, it can be observed that the satisfaction with and adaptation of chatbots as well as voice assistants is falling behind expectations (Cesta and D'aloisi 1999; Fuckner et al. 2014; Krogue 2017; McKnight et al. 2002; Zierau et al. 2020). On the one hand, this is to be expected, as market research has shown that users are generally skeptical of new technologies (Zierau et al. 2020). Development of UASs have been strongly driven by advances in technology rather than emerging from user needs, which has left users' needs and desires not yet fully understood and addressed (Zierau et al. 2020).

In summary, UASs provide many potential benefits for individuals, organizations, and society. To realize these benefits, however, the users' skepticism and hurdles of adaptation must be overcome by developing UASs that understand and address the needs of users (Maedche et al. 2016). UASs should help users navigate an increasingly complex world without introducing even more complexity.

1.2. The Role of User Assistance Systems in Information Systems Research

Maedche et al. (2019) describe UASs as socio-technical systems with three key components. First, the user who wants to fulfill a specific goal. Second, the task needed to achieve the goal. Third, the technology with which the user interacts to perform the tasks. UASs are part of the technology component that helps users perform tasks better (Maedche et al. 2016). The interaction takes place in a specific application context. Figure 1 visualizes this interplay of components of a socio-technical system.

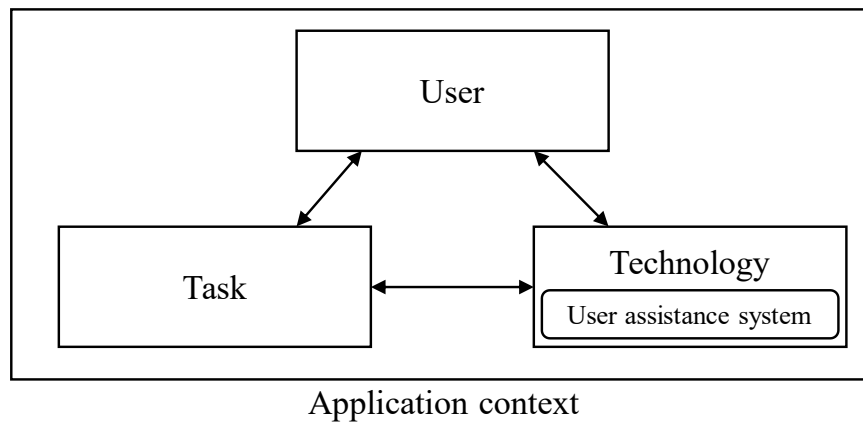


Figure 1: Socio-Technical System (Based on Maedche et al. (2019))

UASs are designed based on the assumption that people and machines each have their own strengths, which, when combined, lead to additional value (Maedche et al. 2019). Machines are well suited for structured and repetitive tasks. Their primary purpose is to collect and process information as well as to draw conclusions. In contrast, humans are better at working on complex and abstract problems as well as drawing conclusions from incomplete data (Maedche et al. 2019). Neglecting the human in this socio-technical system leads to a high vulnerability to errors, making it difficult for machines to respond to unforeseen and complex tasks. If one neglects the UAS, the potential benefits from the interaction and assistance of people will not be realized. Only when the components of the socio-technical system are appropriately aligned with each other do benefits arise (Maedche et al. 2019). If the UAS is too simplistic and does not have the required capabilities, the UAS cannot contribute and the user solely performs the task execution. Vice versa, if the UAS is too complex, the user cannot efficiently use the services provided by the UAS for task execution. Negative consequences could be, for example, confusion, distraction, and additional cognitive load, which in the worst case, can also lead to rejection of the UAS. To gain an advantage from the interaction between the user and the UAS, it is necessary to understand the interaction within the socio-technical system and to align the different components. Thus, knowledge from different research disciplines is required.

IS research can make a valuable contribution to better understanding the interplay between the user, the UAS, and the task (Pfeuffer et al. 2019). IS research traditionally deals with socio-technical phenomena, combining knowledge from other disciplines such as computer science, engineering, psychology, management, and economics (Friedrich et al. 2017). Hence, IS research combines theoretical knowledge and methodological tools from different disciplines to uncover unique findings (Maedche et al. 2019).

In the IS context, user assistance is defined as „a system’s capability to assist users while performing their task by means of human-, task-, and/or context-dependent augmentation of the human–computer interaction“ (Morana et al. 2020, p. 189). Thus, UASs can be defined as applications that enrich IT systems (Maedche et al. 2016) to “bridge the gap between the system’s functionalities and the human’s individual capabilities with the goal of positively influencing task outcomes” (Morana et al. 2020 p. 189). These are software components that generate a specific output through the input and processing of data (Maedche et al. 2019). A key characteristic of UASs is that they do not enforce behavior but rather guide the user during the execution of tasks (Maedche et al. 2016). UASs align system functionalities with individual capabilities to positively influence task performance (Morana et al. 2020). To illustrate the boundaries of this definition, two examples will be examined. Word processing programs (e.g., Microsoft Word) are not viewed as UASs, since they do not assist the user in the execution of the task "writing text", but are a tool that enables the task execution. Spelling support, as a part of most modern word processing programs, can be seen as a simple UAS since it assists the user with additional information and advice during task execution but, at the same time, does not enforce any action on the user. Another example is the previously discussed Whatsapp chatbot for providing Covid-19 information (George et al. 2021a). In this case, the messenger service Whatsapp is not a UAS, as it is only the underlying communication platform and functions as a technological task enabler. The chatbot, however, provides relevant information to the user and represents a UAS.

Different users and tasks also have different requirements for a UAS. Maedche et al. (2016) classify UASs along two dimensions: Interaction and Intelligence. Interaction describes the intensity of the exchange between the user and the UAS and, thus, the capabilities of the UAS to participate cooperatively in task processing (Maedche et al. 2016; Morana et al. 2020). Intelligence describes the ability of the UAS to adapt the assistance to the user's environment and preference by processing context data (Maedche et al. 2016). Intelligence also includes the emotional intelligence of a UAS, i.e., sensing and understanding emotions and responding accordingly (Morana et al. 2020). Based on these dimensions, Maedche et al. (2016) derive four classes of UASs. Figure 2 visualizes these four classes.

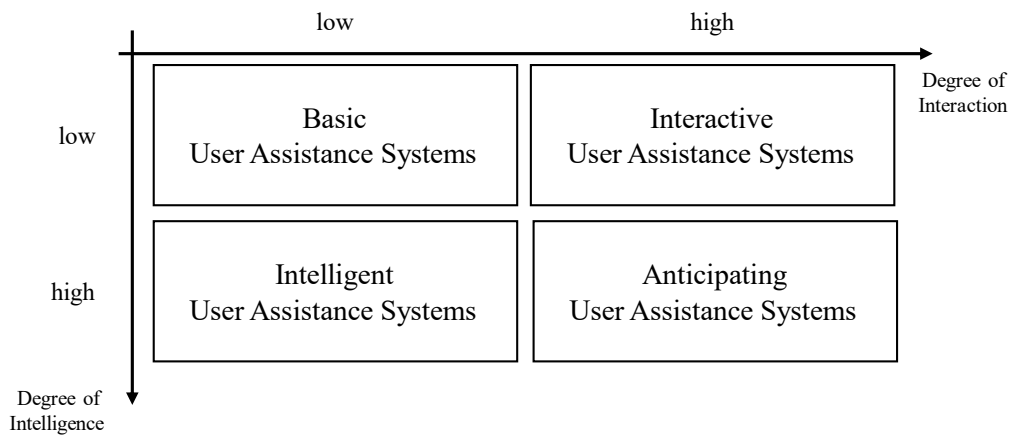


Figure 2: User Assistance Systems Classification (Based on Maedche et al. (2016))

Basic UASs have a low degree of interaction and intelligence. Thus, there is little interaction with the user and the assistance is only slightly adapted to the current application context (Maedche et al. 2016). Typical examples are UASs that mainly provide information and indirect assistance, such as computerized versions of user manuals (Maedche et al. 2016) or spelling assistance in word processing programs.

Interactive UASs have a high degree of interaction but low intelligence. The focus is on cooperation with the user and the increasing participation of the UAS in task execution. Due to the low level of contextual information, interaction is primarily based on fixed behavior patterns, such as giving feedback, reminders, or using gamification elements (Maedche et al. 2016). The interaction can be text-based but also speech-based. Examples of this class of UASs are mobile eco-driving feedback apps (Gimpel et al. 2022) and fitness apps that aim to encourage the user to perform physical activity (Olafsson et al. 2020).

Intelligent UASs collect and process data to adapt the assistance to the current application context. The data can be used to support the user in the task processing itself. For example, recommender systems use the previous purchase behavior of customers to assist them in selecting new products (Maedche et al. 2016). Alternatively, intelligent UASs can adapt to user needs by detecting emotions and responding accordingly. For example, Bromuri et al. (2021) present an approach to an intelligent UAS for improving employee well-being that monitors employee stress levels in real-time so that countermeasures can be initiated if too much stress is experienced.

Anticipating UASs have a high degree of interaction and a high degree of intelligence. They have a detailed understanding of the user as well as the application context and constantly adapt their assistance and anticipate future events. At the same time, they offer sophisticated text-

based or speech-based interfaces (Maedche et al. 2016). One instance of this class of UASs are modern voice assistants that analyze their users' behavior over time, incorporate contextual data, and learn to provide more valuable services. These UASs can, for example, proactively and predictively warn users of future flight delays by automatically searching the user's calendar for flights and calculating the probability of delays based on previous flight data (Carman 2018). This behavior allows anticipating UASs to warn users when booking and suggest other flight alternatives or to assist in calendar management by considering future flight delays when scheduling appointments.

Each of these four classes is differently suited to different application contexts. One might be tempted to assume that all UASs should be designed with a high degree of interaction and a high degree of intelligence to accomplish the most benefits. However, excessive degrees of interactivity and intelligence can also cause disadvantages. Too much interaction requires cognitive resources from the user and can lead to inefficient working conditions and dissatisfaction. On the other hand, the development of intelligent UASs is costlier as more data and more flexible data structures are needed. Thus, too much intelligence would lead to an unnecessary investment of resources. Therefore, it is essential to consider the particular application context when developing UASs and appropriately align the individual elements. Depending on the application context, different types of UASs need to be designed, implemented, and evaluated (Maedche et al. 2016).

1.3. Aim and Outline of this Dissertation

Although there are already several approaches for UASs in IS and related disciplines, more research is still needed (Maedche et al. 2016). The individual interaction between user, task, and UAS in a specific application context leads to a demand for suitable UASs. At the same time, individual, organizational, and societal problems still require more research attention and would benefit from using UASs. Due to the rapid development of technology, even more application domains will emerge in the future (Maedche et al. 2019). Thus, IS scholars call for „a more in-depth, multidisciplinary investigation of user expectations, theories, patterns and guidelines [that] could advance the field.“ (Dhiman et al. 2022, p. 663). It is evident that “as a sociotechnical discipline, the [IS] community is challenged to provide scientifically grounded and practice-relevant answers to the question how [to design] the interplays between users, tasks, and technologies” (Maedche et al. 2019, p. 536). Responding to this call, the research in this dissertation aims to contribute to IS knowledge on UAS design by better understanding the

interaction between UASs and users in different application contexts. Therefore, new theoretical knowledge is developed based on a strong theoretical background from different research disciplines. The new theoretical knowledge is empirically evaluated by collecting and analyzing data to explain and predict user behavior. Researchers and practitioners can use the findings to build user-oriented UASs to reduce adaptation barriers and realize the potential benefits. To generate broad knowledge of UASs, the classification of UASs by Maedche et al. (2016) is used. For each class (basic UASs, interactive UASs, intelligent UASs, and anticipating UASs), this dissertation determines an exemplary application context, presents suitable UAS design features and analyzes the benefits for the users.

First, the dissertation examines the application of basic UASs. This class of UASs has a low level of intelligence and a low level of interaction. This makes them particularly suitable for repetitive tasks in which interaction is not required or desired. The use of basic UASs enables the provision of assistance in a less invasive manner that puts a minimal additional cognitive load on the user. Such an environment is, for example, social media. The primary use of social media in leisure time for entertainment and relaxation causes users to typically minimize the investment of cognitive resources (Moravec et al. 2018). This phenomenon is intensified by the high density of information in social media (Metzger and Flanagin 2013), leading users to engage less with the content and question it less critically (Lewandowsky et al. 2012). This behavior becomes dangerous when people come into contact with false information, so-called fake news. Fake news is information that deliberately contains false information and aims to deceive consumers (Shu et al. 2017). Previous research observed that users tend to simply accept fake news without questioning it (Lewandowsky et al. 2012), which results in negative consequences for the individual and society (Clarke et al. 2021; Laato et al. 2020; Roozenbeek and van der Linden 2019; van der Linden et al. 2017). A basic UAS can provide the necessary assistance to contradict these consequences by warning the users of questionable content and providing them with verified information. For such a fake news intervention to work efficiently, it is important to understand the users' underlying information processing processes. Research shows that the order of information exposure impacts information adaptation, but the findings present contradicting approaches. Research on inoculation theory discovered that showing correct information first can protect users against false information (Azzopardi 2021; Lewandowsky and van der Linden 2021; Loftus 2005). Accordingly, basic UAS design would be beneficial if users first see the intervention and then the fake news story. However, at the same time, research on the misinformation effect shows that new information is more easily

retrieved and can even alter old memory (Abualsaud and Smucker 2019; Crozier and Strange 2018). Thus, basic UAS design would be beneficial if users first see the fake news stories and then the intervention. Hence, it is a mystery in what sequence a basic UAS should present fake news stories and interventions to improve fake news perception. To solve this mystery, this dissertation conducts an experiment and derives conclusions for the design of basic UASs for fake news intervention with warning labels and verified information. The results confirm that the sequence in which fake news stories and corresponding interventions are presented influences the perception of fake news. If the intervention is displayed after the fake news story, the fake news story is perceived as less credible and is less likely to be adopted into memory. The results shed light on the information-processing processes that happen after the interaction between the basic UAS and the user and offer important recommendations for the design of future basic UASs as a tool to improve the response to fake news.

Basic UASs are a promising tool to improve people's individual perception of fake news. However, this approach cannot guarantee protection against fake news. Therefore, in addition to improving individual fake news perception, it is also important to remove fake news from social media platforms as quickly as possible to prevent it from spreading. Social media platforms offer their users the function to report fake news in order to participate in the fight against them. However, anecdotal evidence shows that users do not take this responsibility and do not use this feature. Therefore, this dissertation presents an approach for a basic UAS that motivates users to report fake news by displaying social norms (SNs). SNs indicate what behavior is socially desirable or what behavior others apply in a particular situation (Baumgartner et al. 2011). Previous research showed that SNs are a promising approach to motivate prosocial behavior (Cialdini et al. 1990; Cialdini 2003; Goldstein et al. 2008). However, it is uncertain whether SNs can also have their effect in a semi-anonymous environment such as social media. At the same time, SNs can backfire and promote undesirable behavior (DiFranzo et al. 2018). Thus, it is also uncertain whether SN can be used as a mechanism for a basic UAS to motivate users to report fake news. This dissertation conducts two experiments to better understand users' interaction with a SN-based basic UAS. The results indicate that basic UASs with SNs are a promising tool to motivate social media users to report fake news and thus lay the foundation for more UASs to support users in dealing with fake news.

Second, this dissertation examines the application of interactive UASs. The focus of interactive UASs is on cooperative interaction with the user to achieve a specific goal (Maedche et al.

2016). By implementing regular feedback and gamification elements, interactive UASs are well suited to accompany and support users over a more extended period of time and thus achieve long-term goals. Typical goals include more physical activity, a healthier diet, and less tobacco and alcohol consumption. These so-called SNAP factors (smoking, nutrition, alcohol overconsumption, and physical inactivity) are the four key risk factors for non-communicable diseases, such as heart disease and type 2 diabetes, and, thus, are the leading cause of premature deaths worldwide, with 16 million cases yearly (World Health Organization 2020). Hence, reducing these risk factors by changing individual behavior offers great potential for individuals and society. Research presented different interactive UASs to improve delivering behavior change interventions (BCIs) (Michie et al. 2017; Noorbergen et al. 2019). In this context, anthropomorphic design features can facilitate forming a sociotechnical relationship between the user and the interactive UAS (Kim and Sundar 2012a; Pfeuffer et al. 2019). To benefit from this phenomenon, researchers and practitioners often use digital human representations (DHRs), such as avatars and embodied agents, to deliver BCIs (Aljaroodi et al. 2019; Noorbergen et al. 2019). Many studies demonstrated that DHRs can successfully assist in health behavior change (Bickmore et al. 2013b; Lisetti et al. 2013; Peña et al. 2016; Song et al. 2013). However, there are different social design features to consider when designing DHRs, such as physical representation, language, emotional capabilities, intended social role, or social dynamics during the interaction. Each social design feature has a specific effect on the user's perception and, therefore, a particular impact on the success of BCIs. Thus, it is important to understand how social design features influence the delivery of BCIs. This dissertation reviews and synthesizes existing empirical evidence from previous research on the influence of social design features on BCI success in the SNAP domain. The results highlight the most widely used social design features and analyze their effect on BCIs. These findings provide a valuable contribution to the future design of DHR-based interactive UASs.

Third, this dissertation examines the application of intelligent UASs. Intelligent UASs can adapt to the current application context. They can assist in task execution by processing context-relevant data. Furthermore, they can also adapt to the needs of the users. Intelligent UASs are continuously improving their ability to recognize users' emotions and respond accordingly (André 2021; Schneeberger et al. 2021). Emotional intelligence allows intelligent UASs to stimulate people not only in a task-oriented way but also to assist them emotionally. Therefore, this class of UASs is suitable for use in situations where interpersonal relationships are neglected. Such a setting is the crowd working field (Jäger et al. 2019). Crowd work is “a digital

form of gainful employment that is based on the principles of crowdsourcing in order to orchestrate an undefined mass of people via an open call on IT-facilitated platforms” (Durward et al. 2016, p. 282). In practice, this means that so-called crowd workers select a task on a corresponding platform, process it according to the instructions and are then paid (Jäger et al. 2019). Typical examples of such platforms are Mechanical Turk, Clickworker, and appJobber. Here, crowd workers often work independently and without any relation to other people. However, self-determination theory identifies relatedness as an inherent human psychological need (Deci and Vansteenkiste 2004). Relatedness describes “the universal propensity to interact with, be connected to, and experience caring for other people” (Deci and Vansteenkiste 2004, p. 5). Thus, relatedness substantially impacts the performance and well-being of employees (Deci et al. 2017). Also, traditional leader-employee relationships, which have a positive, stimulating effect on employees, are not developed in a crowd working environment. The more leaders are able to perceive and respond to their employees' emotional states, the more likely they are to create a comfortable and productive work environment (Cavazotte et al. 2012). Intelligent UASs with distinct emotional intelligence have the potential to address the need for an emotion-sensitive leader. This dissertation examines the influence of an intelligent UAS with leadership capabilities and emotional capabilities on the performance and enjoyment of an employee during crowd working. For this purpose, a CA with leadership capabilities and emotional capabilities was developed. The CA detects the user's mood at the beginning of the interaction and adjusts its responses accordingly. The focus of this dissertation is not the detection of the mood but the stimulation of the user during the interaction to investigate the subsequent performance and enjoyment. An experiment was conducted with real crowd workers. The results show that the leadership capabilities of the intelligent UAS lead to an improvement in well-being, but contrary to the expectations, also to a reduction in the performance of the crowd workers. Emotional capabilities reduce the stimulating effect of leadership characteristics. The results contribute to a better understanding of the complex phenomena involved in the interaction with intelligent UASs, while also revealing potential challenges that must be considered in the development of future intelligent UASs.

Fourth, this dissertation examines the application of anticipating UASs. Anticipating UASs are the most technologically advanced class of UASs. Due to their high degree of intelligence and the inclusion of contextual information, they can stimulate users in a task-related and emotional way. Further, the high degree of interaction enables advanced communication interfaces (Knote et al. 2018). Anticipating UASs constantly adapt their assistance to the current application

context and try to incorporate possible future situations (Maedche et al. 2016). These advanced features also change the nature of the interaction. Not only cooperation with the user is possible, but also collaborative teamwork. Anthropomorphic design cues are often used to render the interaction more intuitive and comfortable for the user (André et al. 2019) as well as to initiate a relationship between the user and the UAS (Aldiri et al. 2008; Epley et al. 2007; Pfeuffer et al. 2019). However, anthropomorphic design cues can also lead to overly high expectations. Failing to meet these expectations leads to disappointment and repulsion (Bartneck et al. 2009; Mori 1970). Therefore, to implement a successful collaborative relationship between anticipating UASs and users, it is important to understand how anthropomorphic design cues affect the interaction in collaborative work and what impact they have on decision quality and user satisfaction during collective human-machine decision-making. This dissertation develops a theoretical model based on earlier research to explain the interaction of anthropomorphic anticipating UASs and users. In addition, a suitable experiment procedure is presented to evaluate the theoretical model empirically. The experiment is not conducted, but the experiment design establishes the foundations for future research. This dissertation contributes to a better understanding of the interactions between users and anticipating UASs during collaborative work on a task and provides future research with the necessary tools to advance the field in a meaningful way.

To summarize, this dissertation's theoretical and empirical findings contribute to IS research by creating knowledge on how users interact with UASs and how to improve this interaction to benefit users, organizations, and society. The dissertation provides novel insights across different classes of UASs and different application contexts. Chapter 2 provides the theoretical background on UASs and introduces the methodological approaches used. Chapter 3 examines the application of a basic UAS to understand and improve fake news and intervention perception, as well as to motivate reporting of fake news. Chapter 4 explores the influence of social design features of interactive UASs on the success of BCI in the SNAP domain. Chapter 5 examines the effect of leadership capabilities and emotional capabilities of an intelligent UAS on task performance and enjoyment. Chapter 6 explores the influence of anthropomorphic cues of anticipating UASs on decision quality and user satisfaction during collective human-machine decision-making. Chapter 7 discusses the results concerning the limitations and highlights directions for future research. Figure 3 summarizes the structure of this dissertation and illustrates how each chapter relates to the classification of UASs, according to Maedche et al. (2016).

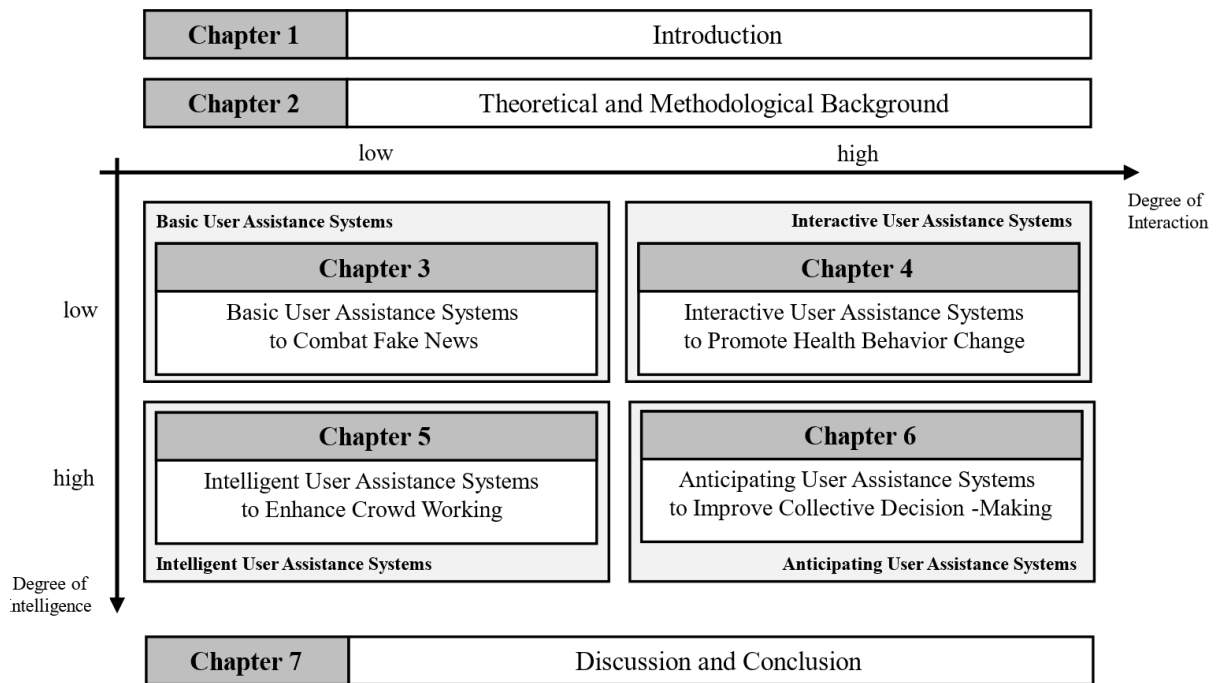


Figure 3: Structure of This Dissertation

During the process of developing this dissertation, parts of Chapters 3 through 6 were published in conference papers and journal papers as part of a regular scholarly discourse or are under consideration for joint publications with coauthors¹. Major parts of Chapter 3 conform with Olenberger et al. (forthcoming) and Gimpel et al. (2021)². Major parts of Chapter 4 conform with Adam et al. (2022)³. Major parts of Chapter 5 conform with Benke et al. (forthcoming). Major parts of Chapter 6 conform with André et al. (2018).

¹ This doctoral thesis follows the “Promotionsordnung der Mathematisch-Naturwissenschaftlich-Technischen Fakultät der Universität Augsburg (vom 21.05.2014, geändert durch Satzung vom 22.07.2015)“ and the “Handreichung des Instituts für Materials Resource Management (MRM) für Doktorandinnen und Doktoranden zur Einbindung von Vorveröffentlichungen in eine monografische Dissertation im Rahmen einer Promotion zum Dr.-Ing. an der Mathematisch-Naturwissenschaftlich-Technischen Fakultät (MNTF) der Universität Augsburg (in der Fassung vom 09.01.2020)“.

² <https://www.tandfonline.com/doi/abs/10.1080/07421222.2021.1870389>

³ <https://aisel.aisnet.org/thci/vol14/iss3/3>

2. Theoretical and Methodological Background

2.1. User Assistance Systems

Assistance has long been a part of IS research and human-computer interaction research (Floyd 1986). In general, assistance describes the performance of an action by one entity for another entity to achieve a specific goal (Dhiman et al. 2022). Dhiman et al. (2022) identify five characteristics of assistance. First, there exists a common goal to be achieved (Dhiman et al. 2022). Second, one entity assists another entity (Dhiman et al. 2022). These entities can be human as well as IT systems and can come in different combinations. Traditionally, a human can assist another human, or an IT system can assist another IT system in achieving its goals. Also, a combination of the different entity types is possible. Thus, a human can also support an IT system when it reaches its limits. Vice versa, an IT system can also support humans or users in various situations. In this case, one speaks of a UAS. Third, an action is performed by one entity to assist another entity in achieving its goal (Dhiman et al. 2022). Fourth, there is a need for assistance (Dhiman et al. 2022). Applied to UASs, this means that the human is not able to achieve the goal on his own or there is a potential for improvement in the execution. Fifth, the assistance should lead to a positive change in goal achievement (Dhiman et al. 2022). This positive change can impact the execution of a task and lead to improved performance, or it can impact the user and contribute to a more positive execution environment (e.g., cognitive relief or emotional support).

UASs are considered agents that support users in task execution. The term agent highlights that UASs are independent entities that can perform specific actions within the application context (Knote et al. 2019). In this context, UASs can interact bilaterally with the user or act as intermediaries between multiple entities (Knote et al. 2019). As a result of the increasing interconnectedness of the world, UASs can interact with other IT systems, e.g., to collect information, disseminate information, or initiate actions. For example, UASs are used in the smart home context to communicate with other devices and automate routine processes (McLean and Osei-Frimpong 2019). UASs are also used as intermediaries in the interaction between human entities. Typical examples are UASs deployed as voice assistants in call centers. Here, the voice assistants can be used to answer the customer's call, record a potential problem, and prepare the information for the employees before putting the customer through (Bae et al. 2005; Qian et al. 2020).

UASs can be utilized at different points of task execution. Wandke (2005) presents six basic application stages for UASs. First, UASs can motivate users and thus initiate task execution (Li et al. 2014; Olafsson et al. 2020; Wandke 2005). Second, it can improve perception during task execution (Wandke 2005). The purpose is to enable the user to perceive information from the environment that the users would otherwise not be able to perceive themselves (Wandke 2005). Possible reasons are that the user is physically unable to do so due to a disability or that another situations demand attention, e.g., when driving a car (Gimpel et al. 2022; Guzman et al. 2018). Third, UASs can prepare relevant information in an appropriate format for users during task execution (Wandke 2005). This stage includes information collection (Kilic et al. 2015) and information preparation (Huang and Rust 2018). Fourth, UASs can provide suggestions for decisions and actions to be taken during task execution (Wandke 2005). For this purpose, the available information is analyzed and recommendations are derived (Maedche et al. 2019). Fifth, UASs can also perform specific actions required for task execution (Wandke 2005). Depending on the application context, task execution can be performed automatically by the UASs or in cooperation with the user (Morana et al. 2017). Sixth, UASs can provide feedback after task execution to improve future task execution.

To systematically explore the research field of UASs, the following section presents different classifications for UASs. There exists a variety of different UASs with different characteristics. Thus, previous research developed different classifications related to UASs focusing on different aspects of a socio-technical system (Müller-Polyzou et al. 2021). Ludwig (2015) classifies UASs into three classes according to their degree of autonomy. *Autonomous UASs* provide simple functions and perform simple tasks automatically, i.e., without user intervention (Ludwig 2015). *UASs with a fixed goal* combine multiple functions to solve more complex problems (Ludwig 2015). These functions can be autonomous, non-autonomous, or a combination of both. The goal is to assist users in performing predefined tasks (Ludwig 2015). *UASs with intention recognition* try to determine the current task from the user's action and provide appropriate actions to support task execution (Ludwig 2015).

The classification of Ludwig (2015) focuses on the autonomy of the UAS and pays less attention to the characteristics of the users. However, the user is an essential element of the socio-technical system. Thus, Romero et al. (2016) distinguish three classes of UAS based on the respective user capabilities that are to be enhanced. *Physical Assistance Systems* assist users in physical activities to enable or enhance movements (Romero et al. 2016). For example, collaborative robots work with users to lift heavy objects (Mark et al. 2019). *Sensorial*

Assistance Systems are used to support the user in perceiving information from the environment for decision-making (Romero et al. 2016). The assistance consists of two components: collecting relevant information and user-oriented information preprocessing (Romero et al. 2016). Exemplary applications include fitness trackers that monitor the user's bio-parameters and notify them of problems (Kroll et al. 2016). *Cognitive Assistance Systems* assist the user in decision-making by using additional information to derive a recommendation (Romero et al. 2016). For example, robo-advisors help users make complex investment decisions (Jung et al. 2018).

In addition to users and UASs, the task itself is also an essential part of the socio-technical system. In this vein, Morana et al. (2017) present a classification that distinguishes UASs according to the distribution of task execution between users and UASs. *Supportive UASs* provide information and functions to the user, but the task execution remains primarily with the user (Morana et al. 2017). *Cooperative UASs* collaborate with the user so that both the UAS and the user take over some parts of task execution (Morana et al. 2017). *Notifying UASs* perform task execution independently and only notify users of progress (Morana et al. 2017).

In this dissertation, the classification of Maedche et al. (2016) is used because it combines the capabilities of a UAS and the interaction with the user during task execution into an integrated concept. This classification is based on four categories: Basic UASs, Interactive UASs, Intelligent UASs, and Anticipating UASs (see section 1.2). The dimensions of intelligence and interactivity serve as the foundation of this classification. In the past, intelligence was considered an ability primarily attributed to living beings, especially humans. Gardner (2000, p. 34) defines intelligence as “a biopsychological potential to process information that can be activated in a cultural setting to solve problems or to create products that are of value in a culture”. Another human-centered definition is provided by Sternberg (1999, p. 293), who defines intelligence as “the ability to achieve success in life, given one's personal standards, within one's sociocultural context”. The core idea is that intelligence is „the ability to learn from experience and adapt to the environment“ (Huang and Rust 2018, p. 156).

With the new capabilities of IT systems and the advances in machine learning and AI, the term intelligence is increasingly being applied to IT systems and UASs. Transferring the concept of intelligence to UASs, intelligent UASs are intelligent IT systems that collect and analyze different forms of data to adapt to the environment and preferences of their users (Maedche et al. 2016). One requirement is the ability to obtain information from the environment. The

information can originate from different sources. First, various sensors can obtain information from the physical world. The availability of a wide variety of sensors, e.g., acoustic, optical, GPS, acceleration, and more (Dhiman et al. 2022), can generate rich data streams. Second, application programming interfaces (APIs) enable the transfer of data between IT systems to obtain information from the digital world. Third, peripherals, microphones, and cameras can enable exchanges with humans to obtain more information. For example, Dhiman et al. (2022) revealed in their work that for exchanges with UASs, peripheral devices such as keyboards, mice, or touchscreens are most commonly used, followed by natural language via text or speech. By analyzing the data from these different sources, UASs can realize different intelligence capabilities. Thus, Huang and Rust (2018) distinguish four different types of intelligence. *Mechanical intelligence* automatically enables UASs to perform routine tasks (Huang and Rust 2018). Routine tasks are characterized by a high degree of structure and little variation. *Analytical intelligence* enables UASs to analyze data to solve problems and recognize patterns (Huang and Rust 2018). For this purpose, machine learning approaches are often used to analyze large amounts of data, derive conclusions, and apply them to new use cases. *Intuitive intelligence* enables UASs to think experience-based and adapt to novel situations (Huang and Rust 2018). To implement intuitive intelligence, deep learning approaches are used to extract fundamental relationships and mechanisms from the data and transfer this knowledge to new fields of application. *Empathic intelligence* enables UASs to recognize users' emotions, react, and show emotions themselves (Huang and Rust 2018). Empathic UASs pretend to understand users' emotions and try to influence them through appropriate signals.

The interaction between humans and UASs is not only characterized by the intelligence of the UAS but also by the degree of interactivity. Kiouisis (2002, p. 372) defines interactivity as „the degree to which a communication technology can create a mediated environment in which participants can communicate (one-to-one, one-to-many, and many-to-many), both synchronously and asynchronously, and participate in reciprocal message exchanges“. Transferring this concept to UASs, interactive UASs are IT systems capable of dynamically responding to the user's actions by collecting, storing, transforming, and displaying information (Benyon 2014; Morana et al. 2017). However, the interaction with the user must be designed differently depending on the application context (Maedche et al. 2019). There are different *directions of interaction*. Interaction can be (i) user-driven by providing input and trigger to the UAS, (ii) UAS-driven by providing information from the environment to the user, or (iii) bidirectional by exchanging information between users and UASs (Knote et al. 2019).

Depending on the application context, different *input types* can provide information to the UAS (Maedche et al. 2019). The input might consist of sensor data, textual input, speech-based input, or vision-based input (Knote et al. 2019). The UAS then processes this input to trigger an *action* or return a *response* (Knote et al. 2019). An action describes an activity performed by the UAS to execute a specific task (Knote et al. 2019). Alternatively or additionally, a text-based, speech-based, visual response with information can be transmitted to the user (Knote et al. 2019). However, not only does the content of the response matter, but also how the content is communicated. Research has shown that when interacting with IT systems, people exhibit social reactions comparable to those in interpersonal interaction (Feine et al. 2019). This implies that IT systems are perceived as social actors (Nass et al. 1994). Therefore, UASs can use different social cues during interaction to trigger emotional, cognitive, or behavioral responses in users (Feine et al. 2019). Fogg (2003) distinguishes five types of social cues: physical (e.g., embodiment or movement), psychological (e.g., personality or humor), language (e.g., the wording of written or spoken messages), social dynamics (e.g., turn taking or praise for good work), social role (e.g., friend or teammate). Depending on the design of the social cues, the interaction can make the resulting outcome more beneficial (Feine et al. 2019; Maedche et al. 2019).

The assistance of humans by UASs can pursue different outcomes. Although the purposes for which UASs are used and the resulting outcomes are very diverse, three basic outcome types can be distinguished. First, the most commonly targeted outcome is increased efficiency (Dhiman et al. 2022; Maedche et al. 2019). The objective is to use UASs to reduce the duration and resources for task execution through partial or complete automation. The focus is on faster and less resource-intensive processing of necessary tasks and not necessarily on improving quality (Dhiman et al. 2022). Examples are smart-home assistants for the automatic execution of actions in households (Mettler et al. 2017) or automated call classification systems in call centers (Liu et al. 2021). Second, UASs are used to increase effectiveness (Maedche et al. 2019), improving the actual quality of task execution. This can be done by providing additional information and services or monitoring during task execution. In this case, the UAS observes the user during task execution, tries to detect deviations as well as errors and offers possible suggestions for improvement (Dhiman et al. 2022). Examples of such UASs are warning systems against fake news (Moravec et al. 2018), driver assistance systems for sustainable driving behavior (Gimpel et al. 2022), robo-advisers for financial decision-making (Jung et al. 2018), and recommender systems for better product selection (McGinty and Smyth 2006).

Third, UASs improve the user experience (Dhiman et al. 2022; Maedche et al. 2019). Here, users are motivated to perform specific tasks, or the physical and psychological load is reduced during task execution to create a hedonic experience (Dhiman et al. 2022). Examples of this class include virtual coaches to promote healthy behaviors (Creed and Beale 2012; Kramer et al. 2020) or stress assistants that monitors stress levels in real-time to trigger countermeasures (Bromuri et al. 2021).

Overall, the interaction between UASs and users is diverse and enables a wide range of application opportunities. These theoretical foundations on the types and characteristics of UASs are important for developing new UASs in new application contexts. This dissertation applies different research methods to better understand the interaction between UASs and users and generate new knowledge about potential design features. The following section provides an overview of the applied methods.

2.2. Research Methods in Information Systems Research

IS research is generally interested in the complex interplay between people, organizations, and technology (Davis and Olson 1988; Hevner et al. 2004; Lee 1999). Thus, IS research seeks “to acquire knowledge and understanding that enable the development and implementation of technology-based solutions” (Hevner et al. 2004, p. 84) to solve important human and business problems. To achieve this objective, IS research draws from different disciplines, such as computer science, engineering, economics, psychology, sociology, or mathematics (Friedrich et al. 2017; Gregor 2006; Hirschheim and Klein 2012). This approach gives IS researchers access to a wide range of theories and models that can be used and combined to better understand as well as design the interaction with IT systems. In addition, IS researchers have access to a broad catalog of research methods that ensure systematic knowledge generation (Friedrich et al. 2017). IS research distinguishes two types of knowledge. First, descriptive knowledge refers to the description of natural, artificial as well as human-related phenomena and the subsequent derivation of natural laws, principles, regularities, patterns, and theories (Gregor and Hevner 2013). Descriptive knowledge is concerned with the natural world and human behavior and is used to better understand how the world works (Baskerville et al. 2018). Second, prescriptive knowledge concerns human-made artifacts (Hevner 2021). It refers to knowledge that describes how to design artifacts that solve a specific problem and thereby improve the world (Gregor and Hevner 2013).

Knowledge generation is based on two complementary but distinct research paradigms: behavioral science and design science (Hevner and Chatterjee 2010). Hevner et al. (2004, p. 76) describe behavioral science as an approach „to develop and justify theories (i.e., principles and laws) that explain or predict organizational and human phenomena surrounding the analysis, design, implementation, management, and use of information systems“. These theories are an important foundation for researchers and practitioners to understand the interactions between people, organizations, and technologies (Hevner and Chatterjee 2010). Since behavioral science is primarily concerned with existing phenomena, behavioral science serves to generate descriptive knowledge (Gregor and Hevner 2013). This descriptive knowledge influences the functional scope, information content, and technical interfaces in the design of IT systems (Hevner et al. 2004). The theoretical mechanisms during human-computer interaction must be understood to implement IT systems in a purposeful way and to increase their effectiveness as well as their efficiency (Hevner and Chatterjee 2010).

In contrast, design science “seeks to create innovations that define the ideas, practices, technical capabilities, and products through which the analysis, design, implementation, management, and use of information systems can be effectively and efficiently accomplished” (Hevner et al. 2004, p. 76). The focus is on the development and evaluation of novel IT artifacts to satisfy previously identified needs (Baskerville et al. 2018). Thus, design science aims to generate prescriptive knowledge (Gregor and Hevner 2013). Behavioral science seeks to understand and explain the relationships and mechanisms in the use of IT systems, while design science research seeks the effective use of IT systems (Hevner et al. 2004). Although behavioral science and design science are two distinct paradigms that address different issues, they should not compete but complement each other (Hevner et al. 2004). The design of IT systems must be based on appropriate descriptive knowledge that explains and predicts human behavior (Hevner and Chatterjee 2010). Descriptive knowledge is an important basis for developing IT systems, as it provides necessary information on natural, social, and human laws, constraints, and capabilities (Baskerville et al. 2018). Vice versa, developing and evaluating IT artifacts generate new data and may uncover new phenomena, which in turn may influence theory (Baskerville et al. 2018; Hevner et al. 2004).

This dissertation aims to better understand the effects and mechanisms during the interaction between humans and UASs in new application contexts and to provide solutions to the current challenges of humans. For this purpose, new models based on previous knowledge are developed and empirically validated. Further, new empirical evidence is obtained through

prototype development to provide new knowledge that informs the development of UASs.

The objective is to generate essential descriptive knowledge about the interactions between humans and UASs in different application contexts and thereby lay the foundations for future research. Therefore, this dissertation uses methodologies from the field of behavioral science.

Behavioral science methodologies can be separated into two distinct categories. First, qualitative research methodologies are primarily used “to understand the perception of an information system by its users, the context within which the system is implemented or developed, and the processes by which changes occur or outcomes are generated” (Kaplan and Maxwell 2005, p. 35). For this purpose, non-numerical qualitative data from observations, interviews, and documents are collected, systematically analyzed, and interpreted to understand how the world works (Kaplan and Maxwell 2005). Typical methods are interviews, case studies, grounded theory studies, or literature reviews (Kaplan and Maxwell 2005; Palvia et al. 2007; Williams 2007). Second, quantitative research aims to explain, confirm, predict, and translate relationships into generalizable theories and models (Williams 2007). Quantitative research methods typically involve hypothesis formation, collection of numerical quantitative data, and subsequent statistical analysis (Williams 2007). Exemplary research methods include experimental research, surveys, and mathematical modeling (Palvia et al. 2007). Qualitative and quantitative research methods can also be combined as a mixed-method approach (Venkatesh et al. 2013).

Using quantitative and qualitative methods, this dissertation builds on previous research and extends it with new findings and knowledge. The main emphasis is on experimental research. Experimental research begins with the development of theories and hypotheses to explain the phenomenon or relationship under investigation (Dennis and Valacich 2001). This is followed by the design of an appropriate study to test the hypotheses (Dennis and Valacich 2001). In experimental research, the influence of independent variables on one or more dependent variables is isolated and tested. Typically, one or more independent variables are divided among different treatments, the outcome is measured in the treatments, and then compared to the previously developed hypothesis (Williams 2007). Only the independent variables must be changed between treatments to exclude the influence of other factors. Because of this rigorous approach, experimental research has a high degree of precision in the measurement, manipulation, and control of variables (McGrath 1981). The primary purpose is to explain relationships between the variables and to test them in order to extend theory (Dennis and Valacich 2001). Therefore, this research method is particularly useful to better understand the

interaction between humans and UASs in new application fields and to validate the effects of prototypical design elements.

Chapter 3 uses three online experiments to apply the applicability of basic UASs in the fight against fake news. Section 3.2 applies an online experiment to better understand in which order a basic UASs should display interventions against fake news and what influence the order has on the perception of fake news. Section 3.3 conducts two online experiments to investigate the influence of basic UASs expressing SNs on the perception and reporting behavior of fake news. Chapter 5 applies an online experiment to investigate the influence of leadership and emotional characteristics of an intelligent UAS on work performance and enjoyment. Chapter 6 presents an approach for an online experiment that will allow future research to explore how anthropomorphic design cues of an anticipating UAS affect collaborative work.

Chapter 4 uses a different research method, as previous research already presents several empirical findings on interactive UASs. Many studies showed that interactive UASs can assist in health behavior change through social design features (Bickmore et al. 2013b; Lisetti et al. 2013; Peña et al. 2016; Song et al. 2013). However, a deeper understanding of how social design features influence health behavior change is needed. Therefore, Chapter 4 uses the method of a systematic literature review. A systematic literature review is used to identify, evaluate, and interpret all relevant literature related to a specific topic, phenomenon, or research area (Kitchenham and Charters 2007). A systematic literature review helps researchers and practitioners understand what body of knowledge already exists, what problems exist, and identifies avenues for new research activities (Levy and Ellis 2006). Therefore, this research method is well suited to identify and analyze the social design features widely used in interactive UASs, thereby generating important knowledge that informs the future design of UASs.

Based on this theoretical and methodological background, the following chapters present different research activities that help to better understand the interaction between humans and UASs and generate new knowledge that informs the design of UASs.

3. Basic User Assistance Systems to Combat Fake News

People strive for knowledge and consume a variety of news to gain it. In recent years, the way news is produced and consumed has changed fundamentally. Most importantly, social media have become essential for gathering and broadcasting information (Shi et al. 2016). Due to a lack of quality control mechanisms, however, these channels are susceptible to misinformation and fake news. For example, following the recent restructuring, the EU warned the social network Twitter that the existing measures against fake news and hate speech are insufficient to comply with the EU's law (Milmo and Rankin 2022). As a result, increasingly negative influence of fake news is evident in social discourse on social and political events (George et al. 2021b; Ng et al. 2021). Fake news is demonstrably false information explicitly used to deceive consumers (Mustafaraj and Metaxas 2017; Shu et al. 2017). Their spreading may lead individuals to believe inaccurate information and has been shown to influence opinions and actions (Allcott and Gentzkow 2017). There are many prominent examples of the real-world impact of misinformation campaigns (Mozur 2018; Roozenbeek and van der Linden 2019). It is to be expected that the spread and influence of fake news will continue to increase and that it will thus become increasingly influential for individuals, businesses, and society.

A central driver for the increase in fake news is the high popularity of social media (Horner et al. 2021). Traditionally, journalists acted as gatekeepers, ensuring the quality of information published in official media outlets. In social media, however, any user can create and disseminate information without adhering to journalistic principles. This leads to high production of unverified information on social media platforms (Rubin et al. 2015). This is particularly problematic because humans' ability to recognize fake news is flawed. In fact, people tend to accept fake news as true without questioning it (Lewandowsky et al. 2012). Furthermore, confirmation bias leads to content that is consistent with existing personal opinions and mental models as more credible (Kim and Dennis 2019). Other factors, such as prior exposure (Pennycook et al. 2018), familiarity (Deng and Chau 2021), laziness (Pennycook and Rand 2019), affective engagement (Maasberg et al. 2018), or mood (Deng and Chau 2021) also have a negative impact on the evaluation of fake news. Due to these vulnerabilities, social media platforms must support users in coping with the abundance of fake news.

UASs have the potential to support users in dealing with fake news in order to reduce the negative consequences. Due to their specific characteristics, basic UASs are particularly suited for use against fake news. Low intelligence is sufficient to assist in the repetitive task of fake

news support. Even though social media content is often unstructured, e.g., texts and images, interaction with information in the social media environment is usually well-structured. Simplified, users repeatedly interact with a series of posts, tweets, or stories that have different content, i.e., text, images, or video. Users can interact with the post (e.g., like, comment, share) or continue to the next post. This repetitive nature allows basic UASs to support users despite low intelligence. In the context of social media, low interaction is an advantage of basic UASs. Social media is used by users for entertainment and relaxation (Moravec et al. 2018). Users want to consume social media content as well as interact with it while minimizing the use of cognitive resources (Moravec et al. 2018). A UAS with a high level of interaction could be potentially irritating as well as cognitively demanding and could therefore be perceived as an irritating distraction. A basic UAS with a low degree of interaction, in contrast, is less invasive and can support the user in dealing with fake news without being too intrusive.

This dissertation explores the use of basic UASs to improve user perceptions and motivate beneficial responses. The structure of this chapter is as follows. Section 3.1 provides the theoretical background. Section 3.2 examines how the order of information presented by a basic UAS affects the perception of fake news. Section 3.3 examines whether a basic UAS equipped with SNs is able to motivate users to actively oppose fake news by reporting the fake news.

3.1. Theoretical Background

Fake News

Fake news in itself is not a new phenomenon. In the past, various groups of people, such as advertisers, political activists, and religious zealots, have used different forms of fake news to influence public opinion or spread propaganda (Mustafaraj and Metaxas 2017; Shu et al. 2017). Two key characteristics make news stories fake news: inauthenticity and deception. On the one hand, fake news contains verifiably false information that is nevertheless deliberately disseminated (Shu et al. 2017). On the other hand, fake news can mislead consumers, whereby incorrect information is interpreted as truthful (Allcott and Gentzkow 2017; Mustafaraj and Metaxas 2017). Fake news appears in the form of text, images, audio files, or videos that are edited or deliberately taken out of context (Mustafaraj and Metaxas 2017; Shu et al. 2017).

Besides the different forms of fake news, there are also different motivations for individuals and organized groups to produce and distribute false information. First, fake news, such as satirical content or hoaxes, serves as a source of entertainment. Second, due to its eye-catching

and sensation-oriented design, fake news is well suited to induce readers to visit the original source (Rubin et al. 2015; Shu et al. 2017) and thereby generate a financial profit through advertising (Allcott and Gentzkow 2017). Third, fake news can influence opinion. The aim may be to create an advantage for the favored political candidate (Allcott and Gentzkow 2017; Balmas 2012; Mustafaraj and Metaxas 2017; Shu et al. 2017; Warzel and Mac 2018) or to influence public opinion on a political or social issue. Therefore, fake news's main danger lies in the fact that people do not recognize the false information and the associated manipulation. This misinformation could lead to an unjustified change in voting behavior or political or social attitudes resulting in a severe negative impact on individuals and society.

Even correcting misinformation does not necessarily change people's beliefs (Flynn et al. 2017; Nyhan and Reifler 2010). Furthermore, even when an individual believes the correction, the misinformation may persist. Any misinformation repetition can have negative consequences, even in the context of refuting it (Greenhill and Oppenheim 2017). This effect is due to familiarity and fluency biases. The more people hear a story, the more familiar it becomes, and the more likely they are to believe it is real (Hasher et al. 1977; Pennycook et al. 2018; Schwarz et al. 2007). As a result, exposure to misinformation can have long-term effects, while corrections may be short-lived. Hence, limiting the spread and impact of fake news is essential.

Fake News in Social Media

Although fake news has already been used in the past, its significance and dissemination have recently gained importance (Shao et al. 2016). Social media, in particular, provides a platform for the effective spread of fake news. First, social media offers a broad audience (Facebook 2019; Twitter International Company 2018). For substantial parts of this audience, social media is an essential source of news (Gottfried and Shearer 2016; Newman et al. 2017), which makes its users highly vulnerable to fake news. Second, the barriers to enter social media are low (Allcott and Gentzkow 2017). For instance, the low effort involved in setting up an account also enables the creation of machine-controlled user accounts, such as social bots (Ferrara et al. 2016). As a result, about 19 million machine-controlled accounts were active on Twitter during the U.S. presidential election 2016, spreading information about the candidates (University of Oxford 2016). Third, transmission speed and reach through social media are higher than ever before. Since exchanging information is an essential part of social media (Kaplan and Haenlein 2010), it is easier to discuss or share both real and fake news stories (Shu et al. 2017). About 25 percent of U.S. adults reported that they shared fake news deliberately or unknowingly

(Barthel et al. 2016). Therefore, information can quickly reach global access virally (Shao et al. 2016; Weedon et al. 2017). The novelty of fake news and the negative feelings they often cause are additional reasons that fake news travels faster than real news (Vosoughi et al. 2018). Simultaneously, decentralized distribution makes it difficult to stop the dissemination of fake news since there is no longer a central origin of spread that can be addressed. Fourth, no authority controls the authenticity of information (Shao et al. 2016). The process of journalistic information verification and the associated gatekeeping functions can no longer be guaranteed because social media allows everyone to participate in news production (Aisch et al. 2016; Park and Smith 2007). Due to the lack of control, validating the authenticity of information now lies with the users (Rubin et al. 2015). Furthermore, the widespread dissemination of fake news may reduce the demand for authentic information (Allcott and Gentzkow 2017).

However, people's ability to decide whether the information is true or false has some natural weaknesses, especially in the social media environment (Lazer et al. 2018; Moravec et al. 2019; Shu et al. 2017). A key factor for validating authenticity is the credibility of information, which is complex and diverse (Petty and Cacioppo 1986; Wathen and Burkell 2002). Different factors, such as source (expertise, knowledge, and trustworthiness), receiver (receptivity for a message), and message characteristics (consistency and clarity), all interact in the assessment of the credibility of information (Wathen and Burkell 2002). Additionally, the beliefs and prior knowledge (Pennycook et al. 2018) of a user are important because these factors can lead to confirmation bias. People prefer to believe information that matches their pre-existing beliefs (Allcott and Gentzkow 2017; Housholder and LaMarre 2014; Kim and Dennis 2019). When they encounter information that challenges their pre-existing beliefs, they experience cognitive dissonance (Festinger 1957; Mills 2019). Resolving this cognitive dissonance takes cognitive effort that people tend to avoid (Simon 1979). Simultaneously, the assessment of other social media participants also impacts the credibility of news articles (Kim et al. 2019). Due to these factors, it is necessary to develop countermeasures that support social media users in handling fake news.

Countermeasures Against Fake News

A body of IS research is emerging that addresses questions around the consumption of fake news (Bernard et al. 2019). The problem of fake news has already been recognized in research and practice and different intervention measures have been presented. Ng et al. (2021) identify three classes of measures. First, community-level measures motivate users to identify fake news

and stop its spread (Ng et al. 2021). Second, account-level measures stop the spread of fake news by limiting the possible actions of malicious user accounts (Ng et al. 2021). This includes blocking accounts (Lazer et al. 2018) or disabling the function to share content (George et al. 2021b). Third, content-level measures highlight information or provide more information about fake news (Ng et al. 2021). Different forms of warning labels (Moravec et al. 2019; Moravec et al. 2022), source ratings (Kim and Dennis 2019), or additional information (Gimpel et al. 2020) have been investigated.

Kim and Dennis (2019) offered evidence that changing how the source of news is presented leads social media users to better reflect on whether to believe articles and perform activities that contribute to their spread. Specifically, both changing the interface to highlight the article's source and providing source ratings by a board of experts positively affected users' cognitive and behavioral processes. In a similar vein, Kim et al. (2019) investigated three different types of source reputation ratings and suggested that these ratings influence the extent to which users believe articles on social media. Moravec et al. (2022) found evidence that the mere fact of asking people to reflect on their background knowledge and rate the truthfulness of a news story made them think more critically about that news story and others. Despite these results, in practice, many social media platform providers are criticized for the lack of countermeasures against fake news (The Editorial Board 2016). Therefore, social media providers tried to develop various countermeasures to prevent the spread of fake news, such as automatically detecting social bots or cyborg users. Another method is to reduce the financial incentives to spread fake news by prohibiting sites from advertising if they have repeatedly shared fake news (Shukla and Lyons 2017). Social media providers often work with third-party fact-checking organizations that manually check suspicious information (Hunt 2017). The social media provider can then flag the corresponding content and display other articles on this topic.

These measures are preferred by social media providers instead of deleting information, even if it is false because they do not want to limit freedom of speech and do not want to take responsibility for deciding what information is considered true (Grady et al. 2021; Levi 2017). Therefore, social media providers present users with information from fact-checking organizations and other sources and let users decide how they evaluate the information (Britt et al. 2019; Grady et al. 2021).

In particular, flagging with warning labels is widely studied (George et al. 2021b). This approach aims to nudge people to engage more deeply with suspicious content rather than

relying on the heuristics described previously. For example, Garrett and Poulsen (2019) and Moravec et al. (2020) show that flagging can reduce belief in and intention to share fake news. However, the overall research results are mixed. Arendt et al. (2019) show in their study that warning labels reduce the perception of the accuracy of fake news for politically left-leaning individuals. However, they could not demonstrate this effect for politically right-leaning individuals. Ross et al. (2018) and Moravec et al. (2019) also fail to show a significant effect of flagging with warning labels on perceptions of fake news. While they demonstrate that flags can increase cognitive activity, they could not observe any effect on perceived believability due to confirmation bias (Moravec et al. 2019). Furthermore, flags can backfire. For example, previous research shows that exposure to flags can increase general skepticism, making even truthful news more likely to be viewed as manipulated (Ross et al. 2018).

Another approach is to provide additional verified information and related articles to help people assess the accuracy and correct possible misconceptions. Qualified institutions, such as fact-checking organizations, can provide verified information manually (Margolin et al. 2018). The use of AI to identify fake news (Della Vedova et al. 2018; Reis et al. 2019) and subsequent extraction of verified information from related articles from credible sources is also an option (Zeng et al. 2021). Gimpel et al. (2020) show that the additional provision of related articles helps users improve the identification of fake news in social media. In contrast, however, previous studies also show that corrections can have a negative effect and make people more likely to believe fake news, as preference-incongruent information can lead people to develop arguments in favor of their own opinion and thereby strengthen their own beliefs (Nyhan and Reifler 2010).

In summary, research does not provide clear evidence of the effectiveness of interventions against fake news. While previous research has contributed greatly to the understanding of the effectiveness of different interventions, there is a need for a deeper understanding of the cognitive processes that influence users' perception.

3.2. The Effect of Intervention Order on Fake News Perception

Previous studies highlight two measures that basic UASs can use to support users in dealing with fake news: First, warning labels provided by a basic UAS can increase user skepticism during exposure to fake news (Lewandowsky et al. 2012). Second, the presentation of additional verified information by a basic UAS can draw the user's attention to the questionable features

of the fake news story (Brashier et al. 2021). However, previous findings are mixed. To develop more effective basic UASs against fake news, it is necessary to understand how measures are effective against fake news. Thus, this section investigates the underlying information processing processes of social media users.

Individuals use mental models for internal personal representation of complex phenomena and events (Lewandowsky et al. 2012). Research suggests that the novelty and timing of information exposure are relevant factors for mental model creation. For example, users generally tend to adopt new information (fake or not) into their mental model (Brashier et al. 2021; Lewandowsky et al. 2012). Due to the large amount of information in social media and its primary use as an entertainment medium, users tend to minimize cognitive effort and not critically question the content they see while browsing. Yet, a correction of previously as correct accepted information is tied to substantial cognitive effort (Lewandowsky et al. 2012). Hence, social media users tend to easily adopt information as true but struggle to correct existing mental models. This emphasizes the importance of the sequence in which individuals see information and how it affects mental model creation.

Theoretical arguments in that regard generally fall into two camps. The first camp revolves around the research on inoculation theory (Lewandowsky and van der Linden 2021; Loftus 2005). Inoculation theory indicates that because information shown first can protect against misinformation, thus, users should be shown interventions (e.g., warnings and additional information) before the fake news story. This approach is meant to make users more skeptical about upcoming false information (Lewandowsky and van der Linden 2021). In contrast, a second camp involves research on the misinformation effect (Crozier and Strange 2018), which postulates that it is more promising to show fake news stories first and then the intervention, as new information is more easily retrieved and can even overwrite previous memory (Crozier and Strange 2018). To ensure the effective use of interventions against fake news in practice, it is necessary to better understand the underlying information processing behavior and the possible consequences, particularly over time and regarding their long-term effectiveness. Yet, to this date, such research contributions are missing. Thus, this section focuses on the following research question: What influence does the order of information presented to social media users as part of a fake news intervention have on the perception of fake news after two weeks? To answer this question, this section proposes a combination of two intervention approaches (warnings and verified information) and investigates whether the order in which they are displayed affects how the information is processed and mental models are built. This work

builds on existing research and derives theoretical hypotheses to answer the research question. Based on a longitudinal online study, this research validates the hypotheses empirically. The findings can be used to equip basic UASs with more effective measures against fake news. This will enable basic UASs to better support users in their perception of fake news.

Hypotheses Development

A significant problem with fake news is that users believe them without recognizing that they are not based on facts. For example, people tend to accept information when no contradictory information is available (Lewandowsky et al. 2012). To understand the processes through which fake news becomes ingrained in a user's memory, mental models need to be understood. „A mental model is an internal representation or simulation that people carry in their minds of how the world works” (Kim et al. 2021, p. 16). Mental models are built dynamically and are based on the person's memory and the current context in which they operate (Johnson-Laird 1983). Because mental models arise from personal experience, they are not necessarily correct or complete (Kim et al. 2021). As a result, contact with fake news often leads to false information becoming part of a person's mental model. Questioning or correcting mental models involves substantial cognitive effort (Lewandowsky et al. 2012).

According to the dual process theory, research differentiates between two cognitive systems used for information assessment. System 1 processes are fast, automatic, effortless, emotionally charged, and heuristic. In contrast, System 2 processes are slower, reflective, controlled, efficient, and rational (Kahneman 2011). However, because human cognitive abilities are limited (Lang 2000), people primarily use System-1-thinking in everyday life (Verplanken and Wood 2006). This is especially true for social media, as the high information density in social media (many posts with different information in brief intervals) would otherwise result in high consumption of cognitive resources (Metzger and Flanagin 2013). This explains the tendency of people not to evaluate fake news in social media properly but to accept them directly in their mental models (Lewandowsky et al. 2012). Questioning them would often simply consume more cognitive resources than users are willing to allocate.

Mindlessly scrolling, i.e., processing in System 1 instead of critically reflecting on the information provided, leads to the phenomenon that false information can be falsely and subconsciously adopted and accepted into mental models (Brashier et al. 2021; Lewandowsky et al. 2012). Fake news flags can increase attention in the short term (Moravec et al. 2018). Yet, users cannot form a correct mental model of a situation by receiving only a warning label and

no other information. Thus, they are forced to form a mental model based on the false information and tag it with a negation tag (Lewandowsky et al. 2012). When they later access their mental model on the issue, they can only retrieve the false information and the negation tag. However, the memory of the negation tag can be lost (Grady et al. 2021; Lewandowsky et al. 2012). As a result, only the false information remains in the mental model and may be seen as the truth for the person. Thus, for a successful and sustainable intervention, it is necessary to provide people with the opportunity to build a correct mental model.

For that, we draw on the results of propaganda research. In this context, the psychological inoculation process describes two key elements: First, a warning label (or flag) increases people's attention and caution and puts them in a state of resistance (Loftus 2005). Second, additional verified information allows for comparison between false and verified information (Lewandowsky and van der Linden 2021). Combining the two elements is crucial because the warning label increases caution and motivates people to actively process the information, i.e., to use System-2-thinking to create a correct mental model based on the verified information.

In addition to the information relevant for information processing, the order of presentation of this information is important. From a cognitive-consistency perspective, new information that contradicts the existing mental model leads to inconsistencies in the mental model (Lewandowsky et al. 2012). To resolve these inconsistencies, it is necessary to compare the existing information and derive new meanings. This process is laborious, requires motivation, and consumes cognitive resources (Lewandowsky et al. 2012). System-1-thinking would solve this conflict by rejecting the new information. In contrast, research on the misinformation effect observes that "participants tended to report incorrect new information rather than the correct information they originally witnessed" (Crozier and Strange 2018, p. 585). This phenomenon indicates that new information has higher importance because it is cognitively more readily available and can overwrite old information (Ayers and Reder 1998; Braun and Loftus 1998). Various studies (Loftus 2005; Wyler and Oswald 2016) demonstrate this effect.

Previous research suggests that when people receive information that does not fit with existing information, a state of cognitive dissonance is created. "Cognitive dissonance occurs when users are presented with two pieces of conflicting information that both cannot be true" (Moravec et al. 2019, p. 10). Cognitive dissonance creates a conflict that requires cognitive effort to resolve (Lewandowsky et al. 2012). Since people use social media primarily for entertainment as well as pleasure (Moravec et al. 2018), and therefore mainly use System-1-

thinking, they try to avoid cognitive effort. Consequently, if users see warning labels and verified information first and then the fake news, they engage less with the differences between the fake news content and the verified information. Since people receive the correct information first, there is no motivation to compare the false content and resolve contradictions. They consume both pieces of information, but the willingness for System-2-thinking is reduced to save cognitive resources.

The alternative would be to display the fake news first, followed by the warning label and verified information. People process the false information first and then see the warning and alternative information. Therefore, we suspect that cognitive dissonance will lead people to evaluate the content as well as the conflict that emerges and thus engage in System-2-thinking. This is a necessary step for them to avoid deliberately adopting false information into their mental model. Otherwise, if people see the verified information first, they have no incentive to resolve the potential conflict. Further, the negation tag can be lost over time (Grady et al. 2021; Lewandowsky et al. 2012), so people can no longer distinguish which information was verified and which was false. However, the cognitive dissonance from receiving fake news first nudges them to invest cognitive effort (Moravec et al. 2019). This, in turn, helps them identify the questionable fake news content, prioritize the verified information, and thereby reduce the believability of the fake news. Therefore, we derive the following hypothesis:

Hypothesis 3.2.1 Providing an intervention with warning labels and verified alternative information directly after fake news reduces believability during exposure compared to providing the intervention before the fake news.

Hypothesis 3.2.2 Providing an intervention with warning labels and verified information directly after fake news reduces the believability of fake news over time compared to providing the intervention before the fake news.

For the success of a fake news intervention, it should not only reduce believability during exposure but also protect against the long-term believability of the fake news. Similar to previous research, a time period of two weeks is considered (Grady et al. 2021). To achieve an effective intervention against fake news, it is necessary to build a correct mental model. At first glance, presenting the intervention after the fake news seems counterintuitive, as previous research shows that it is difficult to correct an existing mental model (Lewandowsky et al. 2012). Yet, previous research also states that the more pronounced a mental model is, the more difficult it is to correct (Walter and Tukachinsky 2020). Since a news article on social media

contains aggregated information (e.g., 280 characters in the case of Twitter), there are usually at most a few seconds between fake news exposure and intervention exposure. We suggest this time is insufficient to build a stable mental model that requires correction, in the sense previous research has referred to. Instead, the information is processed together to build a mental model. The effect of more intensive System-2-thinking triggered by a discrepancy between the fake news and verified information activates higher cognitive resources for conflict resolution. It leads users to engage longer and more intensively with the verified information. Thus, questionable aspects of the fake news stories can be identified. Therefore, we assume that the triggered System-2-thinking (by showing the verified information after the fake news) fosters the adoption of the verified information instead of the fake news. As a result, information processed through System-2-thinking will result in strong mental models that can be retrieved more reliably and for a longer period of time. Therefore, we hypothesize the following:

Hypothesis 3.2.3 Providing an intervention with warnings and verified information after exposure to fake news results in less integration of fake news into the mental model than providing the intervention prior to exposure to fake news.

Hypothesis 3.2.4 Providing an intervention with warnings and verified information after exposure to fake news leads to higher integration of verified information into the mental model than providing the intervention before exposure to fake news.

Methodology

We conducted a longitudinal online study to investigate our hypotheses using two phases. In the first phase, we exposed participants to fake news in combination with different interventions and measured short-term effects. The second phase took place two weeks later to examine effects over time. We conducted a pre-test with nine participants before the first phase of the study. We used the feedback to evaluate the comprehensibility of the instructions and the material as well as the procedure and revised them accordingly.

Participants

For the first phase of our study, we recruited 622 participants (of which 374 participated in the main treatments and 248 in additional control treatments – see section Treatments) from the US via Amazon Mechanical Turk (MTurk) in August 2022. The participants received monetary compensation of \$3.50. To ensure high data quality, we implemented several attention checks. First, a question at the beginning of the survey contained an instruction within the text not to

select any answer to the question. If the question was answered, the study was terminated. Second, participants described how they evaluate news believability in an open-text question. The first author reviewed participants' responses to filter out invalid or nonsensical responses (e.g., 'good study'). Thus, we removed three participants due to invalid responses. Third, we checked which participants were already familiar with the general topics and news stories provided in the survey and were thus unsuitable as participants (because they already possess mental models on the subject). We removed 23 participants who didn't fulfill the first criterion and considered only participants who didn't know any specifics about the general topics of the fake news stories for further analysis. We removed a further 60 responses from participants who stated they already knew at least one of the presented fake news stories. The final sample for the first phase contains 538 participants, consisting of 317 participants in the main treatments and 221 participants in the additional control treatments. The sample mostly comprises employees ($n = 341$) and self-employed individuals ($n = 96$) with an average age of 44.33 years. A total of 52 percent of the participants are female ($n = 279$), 46 percent are male ($n = 247$), and 2 percent prefer not to disclose ($n = 12$). 57 percent of the participants ($n = 304$) would rate themselves politically as rather liberal, and 43 percent ($n = 234$) as rather conservative. More details are provided in Appendix A.1.

The second phase took place two weeks later. The participants received monetary compensation of \$2.00 for their participation. After the exclusion of invalid questionnaires (same mechanisms as in Phase 1), the final sample of Phase 2 contains 493 participants, consisting of 292 participants in the main treatments and 201 in additional control treatments. All participants participated in both Phase 1 and Phase 2. The demographic characteristics of the two phases are comparable.

Task

In our study, we exposed participants to news stories, of which three were fake news. The fake news stories were self-developed to ensure that participants were unfamiliar with the stories and did not have a mental model yet. We checked whether all news was new to the participants and excluded individuals who claimed to know at least one already. After each news article, we measured the perceived believability of the news. The order of the news articles was randomized. After a filler task (lasting about 5 minutes), participants were asked questions regarding the news content and what they believed happened in each story. This step is to measure what information entered the participants' mental models. In Phase 2, participants

were asked again, after two weeks, what they believed happened in each story to measure the impact of the interventions on their mental models over time. Figure 4 provides a simplified overview of the study procedure.

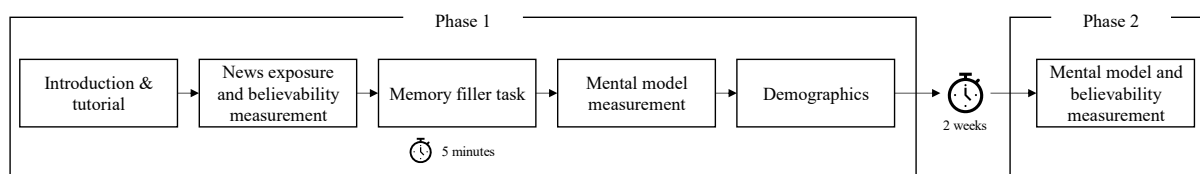


Figure 4: Simplified Overview of the Study Procedure

Treatments

We confront participants with two different fake news interventions and vary their orders to observe differences in the perceived believability and the emerging mental model. To test our hypotheses, we use a between-subjects design with three main treatments:

- (1) Control treatment: No intervention.
- (2) Ex-ante intervention: Participants are first shown a warning label and verified information, and then the fake news story.
- (3) Ex-post intervention: Participants are first shown the fake news story and then the warning label and additional information.

The assignment of participants to the treatments was randomized. Thus, the number of participants per treatment in the first phase varies slightly between 100 and 111 and is between 93 and 101 in the second phase. To exclude the influence of other variables on the results, we test for structural equality of the three treatments. For that, we perform an ANOVA (for age, political orientation, and social media usage) and chi-squared tests (for gender, education, and occupation). At a significance level of 5 percent, no differences between the treatments can be observed. Therefore, we assume structural equality. Appendix A.2 contains the results regarding the structural equality of treatments.

We used two additional control treatments with only warning labels and no verified information to isolate the influence of the verified information. Since this effect is not the focus of this work, we only consider the control treatment and the interventions with verified information (i.e., the three main treatments) in the following. A description of the additional control treatments can be found in Appendix A.3.

Material

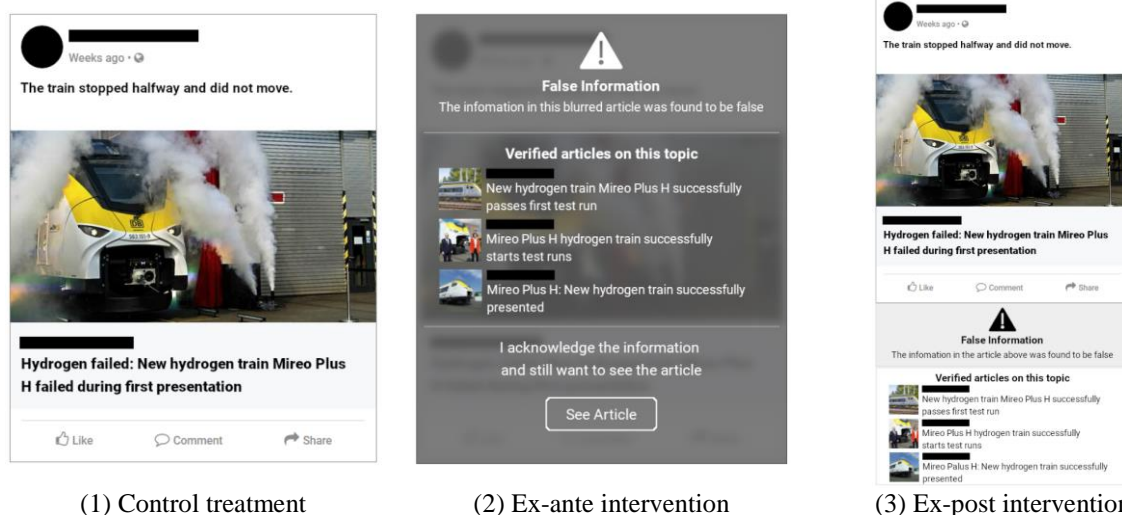
For our study, we developed a total of 21 social media posts. To increase realism and create a familiar environment, we based the design on the most popular social media platform, Facebook (Newman et al. 2017). Posts consisted of source, image, headline, and lead sentence. However, as in previous studies (Kim et al. 2019), we blacked out the source to make people focus on the content. Similarly, we also excluded other confounding factors, such as the number of likes and comments.

The social media posts were split into three categories. In social media, fake news does not occur in isolation but in-between private and real news. Therefore, we first developed three neutral posts, each containing a short video to simulate private posts from the users' social environment. The videos consisted of a video of cats (length: 30 sec), a sunset (length: 30 sec), and a time-lapse of growing plants (length: 19 sec). Second, we created twelve news articles with correct and verified information that originated from authentic news sources. The posts of these two categories serve as the environment in which we distribute fake news. Third, we developed a set of six fake news articles, two in each of the three categories: politically conservative, politically liberal, and without political content. Each participant was randomly shown one conservative, one liberal, and one neutral fake news story. The fake news articles were self-developed to avoid participants' prior knowledge. Additionally, we controlled for participants' prior knowledge of all news articles. Appendix A.4 contains all posts.

In addition to the social media posts, we also developed different intervention elements. First, we designed a warning label following the example of Moravec et al. (2020) and Grady et al. (2021). We used a stop icon to draw the participant's attention to the potential threat of fake news and showed a text explaining that the information in the news article was false. In doing so, we communicate clearly to the participants that this is fake news. Second, we implemented verified information for the events of each fake news article and followed the design of Gimpel et al. (2020). Thus, three more news headlines were displayed with an image, again with the sources blacked out. A caption indicating that the news articles were verified information was also displayed to avoid confusion.

Finally, we varied the order in which the interventions were displayed in accordance with our treatments. For ex-ante interventions, we followed the approach of Grady et al. (2021). The fake news article was blurred, and the intervention was displayed as an overlay. Additionally, to make it clear that the upcoming news article will contain false information, participants had

to actively confirm that they acknowledged the warning label and the verified information and still wanted to see the fake news articles. Without this acknowledgment, participants could not proceed with the study. For the treatments with ex-post interventions, the fake news article was displayed first, and the warning label and verified information were displayed below it. Figure 5 contains an overview of the materials in the different treatments.



(1) Control treatment

(2) Ex-ante intervention

(3) Ex-post intervention

Figure 5: Overview of the Intervention Design in Each Treatment

Independent Variables and Controls

Prior Knowledge and Topic Familiarity: To better understand how different intervention styles affect the mental model formation, we first ensured that participants did not have a mental model by measuring how well participants were familiar with the topics of the study. For this purpose, we adopt a 5-point Likert scale for self-report on internet know-how by Kang et al. (2015) to our content topics. We measured how familiar the participants were with the topics by providing five discrete answers: (1) *'I've never heard of it'*, (2) *'I've heard of this but I don't know what happened / is happening'*, (3) *'I know what happend / is happening but I don't know any specifics'*, (4) *'I generally know about it and know some of the specifics'*, (5) *'I am very well informed about it'*.

Article Familiarity: In addition to familiarity with the topics, we measured whether participants already knew the specific news articles to rule out the presence of a mental model. For this purpose, we use the scale by Grady et al. (2021) to track the recognition of news articles, with the answers *'yes'* and *'no'* available.

Confirmation Bias: To measure confirmation bias, we use the approach suggested by Kim and Dennis (2019) consisting of importance and position. We measure the perceived importance of

the news article for the participant using a 7-point item (*'Do you find the issue described in the article important?'*; 1: *'Not at all'*, 7: *'Extremely'*) and multiply it by the participant's position (-3: *'Extremely negative'*, +3: *'Extremely positive'*). The resulting scores vary between -21 and +21 (Kim and Dennis 2019).

Political Orientation: We assessed the political orientation using the corresponding scale by Wohl and Branscombe (2008) based on an 8-point item (*'Politically, I would say I am ...'*, 1: *'Very liberal'*, 8: *'Very conservative'*).

Social Media Usage: To measure the regularity of the use of social media, we use a 6-point item scale (*'How often do you use online social networks (e.g., Facebook, Twitter, or Instagram) in your everyday life?'*, 1: *'Never'*, 6: *'Several times a day'*).

Dependent Variables

Believability: To measure the perceived believability of each news story, we use the established 7-point Likert scale by Kim and Dennis (2019) consisting of three items.

Mental Model: A mental model can be a complex and multifaceted construct. In our study, however, we control that participants do not yet possess a mental model. To measure what information shown during the study participants maintained in their mental model, we use a simplified recognition test analogous to the one used to study the misinformation effect (Takarangi et al. 2006). For this purpose, we provide the participant with a generic statement that refers to the content of the news articles. To increase realism, we formulated them like questions from another person that simulate a conversation on the topic (e.g., *'I heard that a hydrogen train was recently introduced. Do you know how the presentation went?'*). We provided five response options in random order: (1) with the content from the verified information, (2) with the content from the fake news story, (3) and (4) with new fictional content that was not previously shown, and (5) a response that no credible information has been received. Thus, we can distinguish whether participants adopt the fake news story, the verified information, or none of them into their mental models.

Article Recognition: To measure which parts of an article and intervention participants remembered after exposure, we developed a custom 5-point Likert scale to capture how well participants remembered the article, the warning label, and the verified information (e.g., *'Do you remember seeing a warning message (during the first session) stating the information in*

this article was found to be false?', 1: 'Definitely no', 5: 'Definitely yes'). In Appendix A.5, we report all scales.

Procedures

Phase 1: First, in Phase 1a, we informed the participants that they were participating in a study on social media behavior. Fake news was deliberately not mentioned. Then, we asked the participants about their previous knowledge of the topics in the news articles. The participants then went through a tutorial in which we explained the individual components of a social media post. Depending on the treatment, we provided participants with an explanation of the included intervention elements. The tutorial's goal is to ensure that participants understand what types of information they will see during the study.

The second step was news exposure. Here, we showed participants nine news articles in random order. These consist of six real news articles and three fake news articles. After every third news article, we showed one of the short videos. Depending on the treatment, the posts contained interventions. After seeing each post, participants were asked about the believability of the article and their potential confirmation bias.

The third step was a filler task. Participants had to solve a 7x4 memory game in a maximum of 5 minutes. To increase the cognitive effort of the participants, the goal was to find the corresponding solution for simple mathematical equations (e.g., '25-8' and '17'). After 5 minutes, the participants could continue without finishing the task. Such filler tasks are used in research to empty participants' short-term memory and simulate time passage (Walter and Tukachinsky 2020). This procedure allows us to observe the short-term effects of the interventions.

Fourth, in Phase 1b, we measured the mental model for three real and three fake news stories. To ensure that participants were further motivated to think about the news articles' content, they were informed that they could receive a bonus of \$0.10 per correct answer. Additionally, we ensured that participants would not receive the information regarding their achieved individual bonus until the end of the second phase (regardless of participation) so that they could not draw any conclusions regarding the news articles.

Fifth, article recognition was asked to check how the believability has changed and whether the participants can remember the original articles and intervention. For this purpose, we showed them the three fake news articles again and three real news articles in random order. Here, the

articles were shown in a reduced form, i.e., without pictures, so participants would focus only on the content and not recognize the pictures. Furthermore, we added six new real news articles. As in the second step, we asked participants about the believability of each news article. Thus, they were asked if they remembered the exact article, the warning label, or verified information. Phase 1 ended with a short survey on demographic characteristics and a debriefing.

Phase 2: We follow the example of Grady et al. (2021) using a two-week period to observe long-term changes in the effect of the interventions. The participants were informed via email about possible participation before the start of the second phase. In Phase 2, the renewed measurement of the mental model began directly after a short introduction and was followed by the article recognition test. The procedure was identical to the first phase. Figure 6 summarizes the study design.

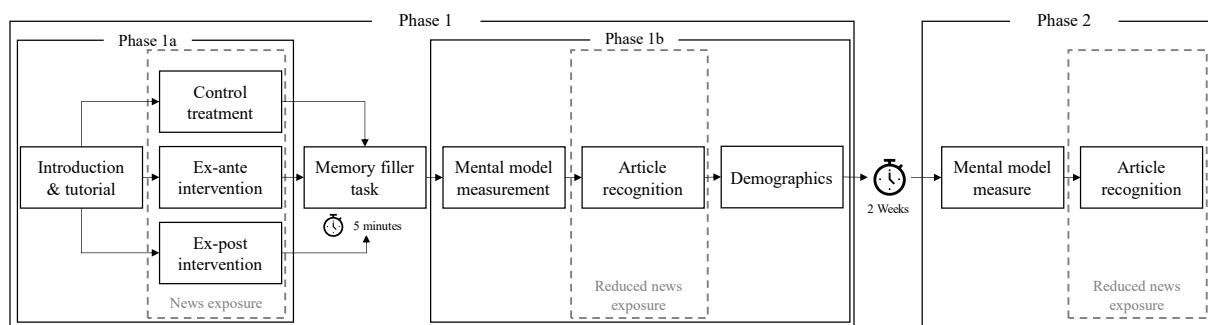
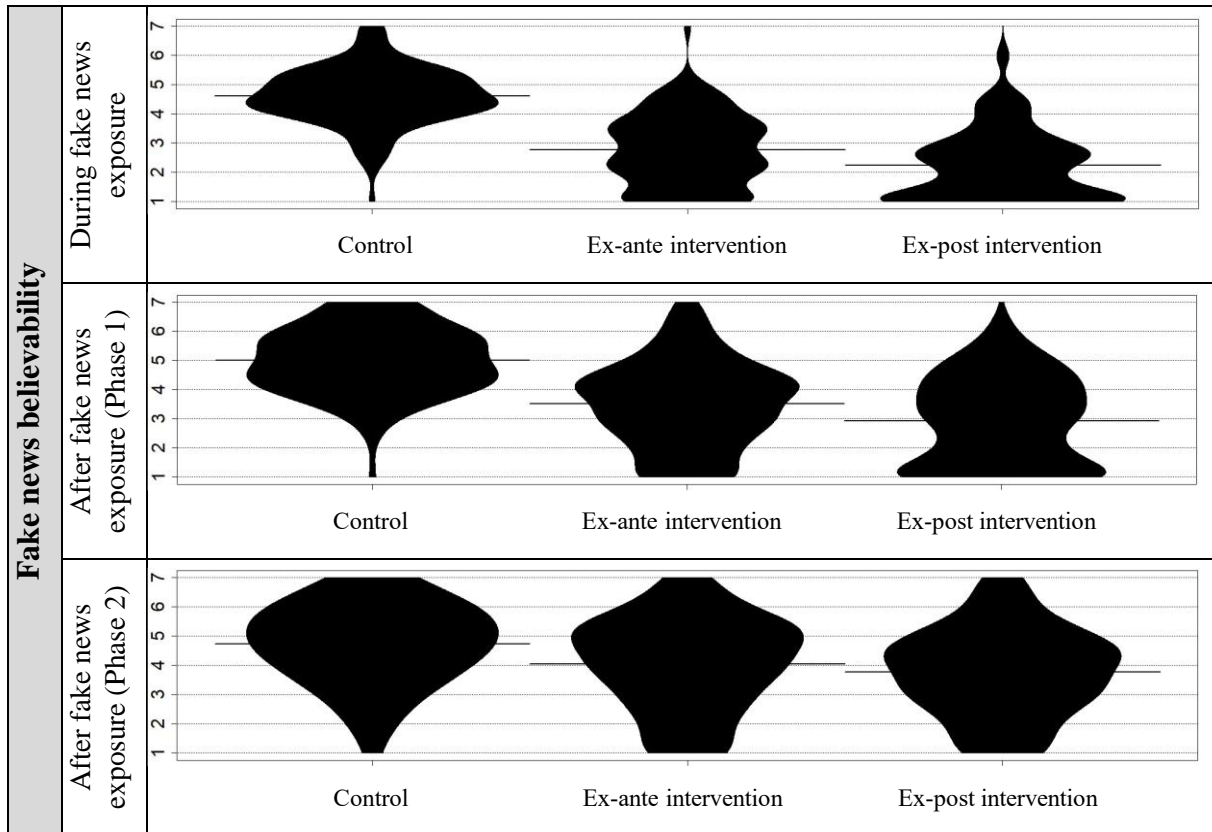


Figure 6: Summary of the Procedure of the Study

Results

Descriptive Data Analysis

An initial descriptive analysis shows a distinct difference between the treatments. Figure 7 contains a comparison of the mean believability of the fake news per treatment for each participant at different points in time. The intervention groups have a lower average perceived believability of fake news. The descriptive data suggest that this difference is particularly pronounced during fake news exposure and decreases over time. Regarding the order of interventions, the ex-post intervention consistently has the lowest believability. The results show that this effect reduces over time. Appendix A.6 contains a detailed descriptive analysis and Appendix A.7 contains a descriptive overview of the data.

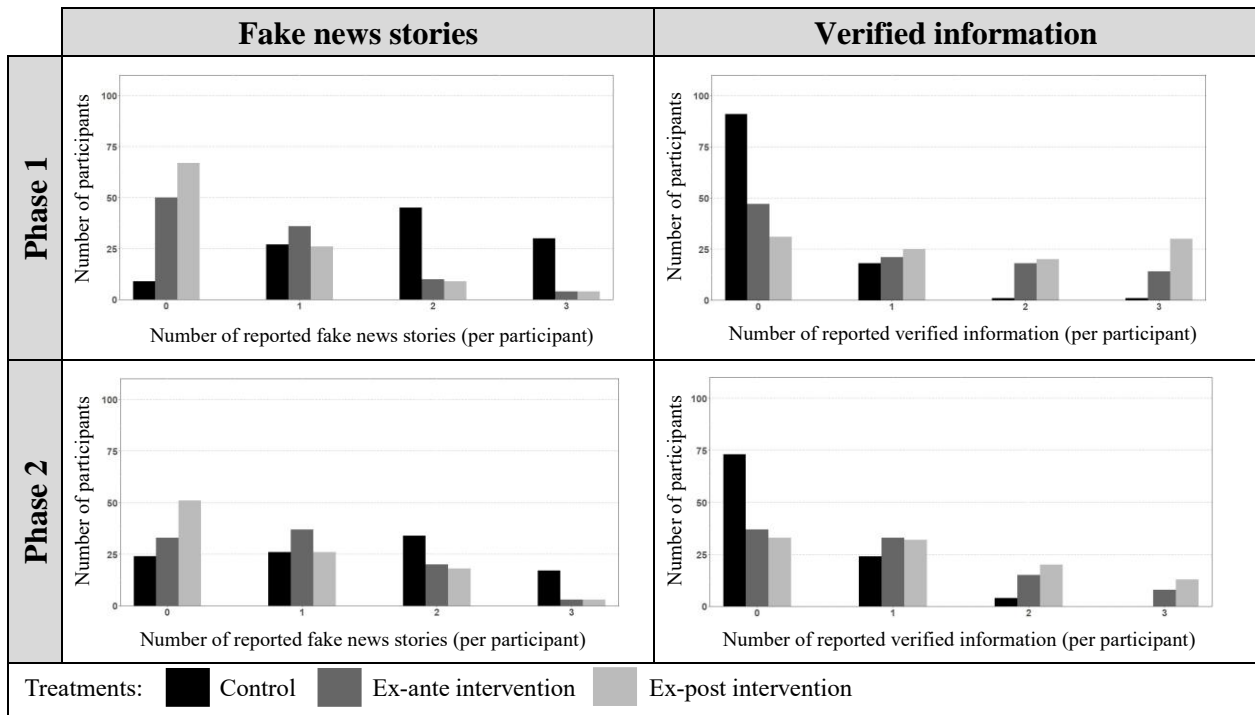


Note: Mean value of the distribution is shown as a horizontal line.

1 - low believability, 7 - high believability

Figure 7: Comparison of the Mean Believability of Fake News

Figure 8 visualizes the results of the measurement of the mental models. For this purpose, we consider the number of fake news stories that were adopted into the mental models for each participant. In other words, the number of fake news stories the participants selected as answers in a conversation on the topic. This yields a value between zero (no fake news story was adopted into the mental model) and three (all fake news stories were adopted into the mental model). The descriptive results reveal that participants with interventions reported less fake news. Conversely, several participants of the control treatment more often reported two or three fake news stories. Furthermore, the ex-post intervention treatment shows the best results. Comparing these results with phase two, we see that the difference is getting smaller, but we observe the same order. In summation, examining how many participants have taken over the verified information into their mental models, we see both in Phase 1 and Phase 2 that the ex-post intervention treatment yields the highest reporting of the verified information.



Note: Numbers vary from zero (the participant adapted no fake news story/verified information into the mental model) to three (the participant adapted all fake news stories/verified information into the mental model).

Figure 8: Mental Model Measurement per Treatment

Hypothesis Testing

The Effect of Intervention Order on Believability: To test the effect of intervention order on believability during fake news exposure (hypothesis 3.2.1) and after fake news exposure (hypothesis 3.2.2), we compare the different treatments by believability. The data analysis using the Shapiro-Wilk test and the Anderson-Darling test (Razali and Yap 2011) suggests that the data are not normally distributed. Therefore, we use a one-sided nonparametric Wilcoxon-Mann-Whitney pairwise comparison. Table 1 contains an overview of the results. Appendix A.8 contains a detailed report of the results.

To test hypothesis 3.2.1, we compare the average believability of the three fake news articles during exposure. The average believability with ex-post intervention is lower than with ex-ante intervention and in the control treatment. Compared to the control treatment, both the ex-post intervention (p-value < 0.001) and the ex-ante intervention (p-value < 0.001) show a significant positive effect. Further, we see a significantly stronger positive effect on believability of the ex-post intervention than the ex-ante intervention (p-value < 0.001). Therefore, the results support hypothesis 3.2.1.

Believability		Mean	Standard Deviation	p-values (Wilcoxon-Mann-Whitney)		
				Control	Ex-ante intervention	Ex-post intervention
Phase 1a (H 3.2.1)	Control	4.621	0.999	-	< 0.001 ***	< 0.001 ***
	Ex-ante intervention	2.778	1.247		-	< 0.001 ***
	Ex-post intervention	2.235	1.198			-
Phase 1b (H 3.2.2)	Control	5.015	1.152	-	< 0.001 ***	< 0.001 ***
	Ex-ante intervention	3.521	1.486		-	0.005 **
	Ex-post intervention	2.932	1.529			-
Phase 2 (H 3.2.2)	Control	4.735	1.391	-	0.002 **	< 0.001 ***
	Ex-ante intervention	4.050	1.604		-	0.070 ⁺
	Ex-post intervention	3.766	1.521			-

Note: Results based on a one-sided Wilcoxon-Mann-Whitney pairwise comparison.

⁺ $p < 0.1$, * $p < .05$, ** $p < .01$, *** $p < 0.001$

Table 1: Results of the Believability Comparison

To examine the effects of the intervention order over time, we look at the average perceived believability in Phase 1b and Phase 2. The results from Phase 1b show that both the ex-post intervention and the ex-ante intervention led to significantly lower believability (p -value < 0.001 for both tests) compared to the control treatment. At the same time, the data show that the ex-post intervention has a significantly larger positive effect on the believability of fake news than the ex-ante intervention (p -value = 0.005). However, this effect is reduced over time. Results in Phase 2 still show a reduction in believability by the ex-post and the ex-ante intervention compared to the control treatment (p -value = 0.002 and p -value < 0.001). However, two weeks after exposure, no statistical difference between the interventions can be observed (p -value = 0.070). Thus, we can only partially support hypothesis 3.2.2.

The Effect of Intervention Order on Mental Model Formation: To test the effect of the order of intervention on the mental model, first, we compare how many fake news participants adapted into their mental model. Second, we compare how many verified information stories participants adapted into their mental models. The Shapiro-Wilk test and the Anderson-Darling test (Razali and Yap 2011) reveal that the data are not normally distributed. Thus, we use a one-

sided nonparametric Wilcoxon-Mann-Whitney pairwise comparison. Table 2 and Table 3 summarize the results.

Number of fake news adopted into mental models		Mean	Standard Deviation	p-value (Wilcoxon-Mann-Whitney)		
				Control	Ex-ante intervention	Ex-post intervention
Phase 1b (H 3.2.3)	Control	1.865	0.909	-	< 0.001 ***	< 0.001 ***
	Ex-ante intervention	0.680	0.815		-	0.045 *
	Ex-post intervention	0.528	0.807			-
Phase 2 (H 3.2.3)	Control	1.436	1.034	-	< 0.001 ***	< 0.001 ***
	Ex-ante intervention	0.925	0.837		-	0.032 *
	Ex-post intervention	0.725	0.871			-

Note: Results based on a one-sided Wilcoxon-Mann-Whitney pairwise comparison.

⁺ $p < 0.1$, * $p < .05$, ** $p < .01$, *** $p < 0.001$

Table 2: Results of the Adaptation of Fake News

To test the effect of the ex-post intervention on fake news integration into the mental model, we consider the number of fake news stories participants adopted into their mental models in Phase 1b. The evaluation of the Wilcoxon-Mann-Whitney pairwise comparison with the control treatment shows that both treatments lead to less fake news in the mental model of a participant (p -value < 0.001 for both tests). A direct comparison of the interventions also shows a significant short-term reduction of fake news in the mental model in the ex-post intervention treatment (p -value = 0.045) compared to the ex-ante intervention.

Further, we consider the results from two weeks later (Phase 2). Both interventions continued to have a significant positive effect (p -value < 0.001 for both tests) compared to the control treatment. A direct comparison of the interventions shows a significant positive effect of the ex-post intervention on the integration of fake news stories into the mental model over time (p -value = 0.032). Based on the results, we can support hypothesis 3.2.3.

Number of verified information adopted into mental models		Mean	Standard Deviation	p-values (Wilcoxon-Mann-Whitney)		
				Control	Ex-ante intervention	Ex-post intervention
Phase 1b (H 3.2.4)	Control	0.207	0.488	-	< 0.001 ***	< 0.001 ***
	Ex-ante intervention	0.990	1.105		-	0.002 **
	Ex-post intervention	1.462	1.189			-
Phase 2 (H 3.2.4)	Control	0.317	0.546	-	< 0.001 ***	< 0.001 ***
	Ex-ante intervention	0.935	0.953		-	0.097 +
	Ex-post intervention	1.133	1.032			-

Note: Results based on a one-sided Wilcoxon-Mann-Whitney pairwise comparison. The control treatment contained no verified information.

+ $p < 0.1$, * $p < .05$, ** $p < .01$, *** $p < 0.001$

Table 3: Results of the Adaptation of Verified Information

To examine the effect of the ex-post intervention on the adoption of the verified information into the mental model, we consider the number of verified information stories participants recalled in their mental models. We observe in Phase 1b that participants in the ex-post intervention treatment were more often able to recall the verified information in the mental model than in the ex-ante intervention treatment. Thus, a Wilcoxon-Mann-Whitney pairwise comparison shows a statistically significant difference (p-value = 0.002). However, the effect is reduced over time. In Phase 2, we see more verified information adopted into the mental model in the ex-post intervention treatment compared to the ex-ante intervention. However, the effect is not statistically significant (p-value = 0.097). Thus, we can only partially support hypothesis 3.2.4. Table 4 summarizes all empirical results.

Theoretical Hypotheses		Empirical Results
H 3.2.1	Ex-post intervention reduces believability of fake news compared to an ex-ante intervention (during exposure)	Supported
H 3.2.2	Ex-post intervention reduces believability of fake news compared to an ex-ante intervention (over time)	Partially supported
H 3.2.3	Ex-post intervention reduces integration of fake news into mental model compared to an ex-ante intervention (over time)	Supported
H 3.2.4	Ex-post intervention increases integration of verified information into mental model compared to an ex-ante intervention (over time)	Partially supported

Table 4: Overview of Hypotheses and Empirical Results

Discussion

The empirical results support our hypothesis that the intervention with warning and verified information after the fake news exposure (ex-post) reduces believability more than the intervention before the fake news exposure (ex-ante), even though the same information is displayed. Per our hypothesis, users engage more with the information to correct the false information they first received. The cognitive dissonance resulting from the two contradictory pieces of information (Moravec et al. 2019) leads to a conflict. If people receive the verified information first, they are not motivated to resolve the conflict and engage with the content.

Yet, when people receive the fake news articles first, they are more likely to resolve this conflict by activating System-2-thinking and recognizing the questionable features of the fake news stories. In addition to the positive effect on believability during fake news exposure, we also observe a lasting effect over time. We find evidence that intervention after fake news exposure results in users adopting the fake news stories less into their mental model. Further, the effect becomes more pronounced over time. After two weeks, we see more clearly that the participants incorporate less fake news into their mental models when the intervention takes place after the fake news exposure. This can also be attributed to increased System-2-thinking.

Second, we observe that participants adopt more verified information into their mental model after fake news exposure when the intervention appears after the fake news story. Because participants engage with the information more intensively when the intervention occurs after fake news exposure, they can recall the correct information more easily later (Azzopardi 2021). However, this effect is weaker over time. Although we observe a positive effect, it is no longer significant after two weeks. Admittedly, the challenge for the participants in Phase 2 of the experiment was high. The participants had to recall externally provided information they had no personal connection with and that was shown to them two weeks earlier as part of a single social media post. Thus, it seems plausible that the gap narrows over time without repeated exposure.

Third and because of the differences in the mental models, we also see reduced believability of fake news stories in the short term. However, this effect is no longer significant after two weeks. It may seem counterintuitive that we see no significant differences in perceived believability, although participants adopt less fake news in their mental model. An explanation could be that when assessing believability, participants first have to activate their corresponding mental model and compare it with the information shown. After five minutes, the activation of memory

and mental model building needs less cognitive resources. After two weeks, however, participants have to invest more cognitive resources to reconstruct the mental model. We suspect that when participants are explicitly asked about their mental model and are monetarily motivated to do so, they are more willing to invest the necessary cognitive resources. We suggest this may be a reason for the difference in the results between believability and the mental model.

In summary, the order in which the intervention is displayed impacts the perception of fake news. The use of intervention directly after the fake news exposure leads to people believing the fake news less, adopting the false information less in their mental model, and continuing to believe it less in the period shortly after the exposure. At the same time, we also observe that while fewer fake news stories are retained in the mental model after a longer period of time, the believability assessment does not correspond in a similar fashion. Appendix A.9 additionally contains a comparison of interventions with only a warning label.

Contribution to Theory

First, we provide theoretical and empirical support that the order of information presented can impact an intervention's effectiveness against fake news. We show that the order in which information is processed can change the perceptions of fake news both in the short term and over time. By looking at different points in time, we contribute to longitudinal empirical evidence for the design and study of interventions (Grady et al. 2021). We extend existing knowledge on interventions by showing that it is important to consider not only which information needs to be presented but also in what order.

Second, we provide an explanation for the cognitive process involved in exposure to fake news and interventions. In particular, we provide further insights into how people form mental models, what influence the order of interventions has, and what the consequences are for perceived believability. Thus, we provide further evidence that using warnings alone is insufficient to address fake news in the long run (Moravec et al. 2019; Ross et al. 2018). Instead, fostering that users build a correct mental model to protect them from fake news over time is crucial for platform providers.

Our findings on the cognitive processes involved in the perception of fake news and the influence of intervention order add to the emergent body of IS research on fake news measures. Thus, we answer the call for more research on the cognitive processes involved in dealing with

fake news and longitudinal empirical evidence (Grady et al. 2021). Our findings can serve as a starting point to develop new interventions and improve existing approaches.

Practical Implications

Our findings can assist practitioners and social media platform providers in developing interventions against fake news. First, only warning labels are insufficient to achieve a long-lasting effect in the fight against fake news. In addition to a warning, verified information should be provided to the user to help build a mental model of the situation. Second, encourage developers to incorporate mental processes into decision-making when designing interventions against fake news. In a fast-paced environment with high information density, participants try to save cognitive resources through System-1-thinking. Therefore, additional information should be displayed after the fake news to motivate users to invest more cognitive resources and engage more with the content. Third, in line with previous research (Grady et al. 2021), we observe a decrease in intervention effectiveness. Therefore, interventions need to be used regularly for continued support of social media users.

Limitations and Future Research

Our results are subject to limitations. The design of our study materials is one of them. Our fake news and verified information were designed to show a clear contrast between each other. Often, the line between fake news and true information is not so obvious, and verified information is not always available. In a real-world setting, the differences may thus be smaller than the ones observed in this study. Future research could verify the findings in real-world settings.

Furthermore, there were subtle differences in the design of the treatments. In the treatment with the interventions before the fake news (ex-ante), participants had to read and actively confirm the intervention information and saw only the fake news after confirming. The intervention appears directly below the fake news article in the ex-post treatment. We cannot eliminate the possibility that the simultaneous exposure to fake news and intervention impacted the results, as participants were able to compare the differences more carefully.

The duration of the study also limits the generalizability of the results over time. The measurements were carried out approximately five minutes and two weeks after the exposure to fake news. In the first phase, this allowed us to simulate the critical moments when fake news is just spreading. In the second phase, we simulated the situation when users reencounter the

fake news after a more extended period of time. However, we cannot say whether the effects persist after longer than two weeks. Yet, two weeks are generally considered an adequate time frame for such studies (Grady et al. 2021). Further research could validate the longevity of our results and could examine a more extended period. Our results show that the effect of interventions is reduced over the two weeks. Yet, it is unclear whether the effect continues to decline over time, is lost, or reaches a plateau.

Lastly, a mental model is a complex construct. In our study, we focus on creating a mental model that does not yet exist to analyze information processing. We control for this in our study. In everyday life, however, users may encounter fake news about a topic for which a mental model already exists. For example, previous research has shown that phenomena such as confirmation biases (Kim and Dennis 2019) make it difficult to correct an existing mental model (Lewandowsky et al. 2012). Future research could shed light on the corresponding information processing to better understand how fake news operates.

Conclusion

This section investigates the influence of the order of basic UAS intervention regarding fake news. Due to their non-inversive nature, basic UASs are particularly suitable for countering fake news in social media. However, basic UAS design needs to consider how users process information to ensure the long-term effectiveness of such interventions. This work theoretically derives hypotheses and validates them with a longitudinal study. The results show that changing the order (of whether the fake news story or the intervention is displayed first) significantly impacts the intervention's success. The findings reveal that users are stimulated to resolve cognitive dissonance when a basic UAS displays additional information after the fake news article. This leads to System-2-thinking that requires more cognitive effort and creates a clearer and more correct mental model. This work contributes to research by advancing the understanding of information processing of fake news. At the same time, the results can be used in practice in the fight against fake news to design new basic UASs. This may reduce the negative effect of fake news on individuals and society.

3.3. The Effect of Social Norms on Fake News Reporting

To improve user perception during fake news exposure is an important first step in the fight against fake news. However, it is also important to remove fake news from social media platforms and stop its spread. Otherwise, effects such as the false consensus effect (Schulz et

al. 2020), illusory truth effect (Hassan and Barber 2021), or availability bias (Spohr 2017) might occur, which make fake news more credible due to frequent and repeated exposure. Social media platform providers are under increasing pressure to take more proactive actions to stop the spread of fake news. Many social media platforms provide their users with the option to report fake news so that the fake news can be professionally checked, flagged, or even removed. However, anecdotal evidence suggests that users only sporadically report news stories as fake. From a theoretical perspective, previous research on the bystander effect shows that people's willingness to help is reduced the more other people are present (Fischer et al. 2011). This effect applies not only in the case of physical presence (Latané and Darley 1970; Schmehl and Lytvynenko 2017) but also in an online environment (Fischer et al. 2011). Following the bystander intervention model, a bystander must follow five steps: noticing and appraising an event as an emergency, recognizing their own responsibility, deciding to intervene, and finally intervening (Kaplan and Haenlein 2010). These steps can be affected by the presence of other people. For instance, a larger group size leads to reduced response rates and willingness to help in online communities (Voelpel et al. 2008) or during e-mail communication (Barron and Yechiam 2002; Blair et al. 2005). In social media, where many users see the same content, this could lead to a diffusion of responsibility (Voelpel et al. 2008), where the individual does not take responsibility for reporting fake news. The user not feeling accountable is another factor contributing to this effect. Within a group of people, an individual evaluates its accountability as lower because another individual may be more accountable to intervene in a specific situation. Making accountability transparent may reduce this effect. However, transparency can also negatively affect when users feel that their intervention is evaluated by others and, therefore, do not dare to take action (DiFranzo et al. 2018). For these reasons, and because of the importance of the other countermeasures, it is necessary to show social media users that reporting is desirable. The question arises, how to motivate users to report fake news?

Social psychology suggests that SNs strongly influence behavior. They might be a solution to the riddle. SNs can steer behavior in a helpful direction for individuals and society (Cialdini et al. 1990; Cialdini 2003; Cialdini et al. 2006). There are two types of SNs: Injunctive SNs describe which behavior most people approve or disapprove of. Descriptive SNs refer to what other individuals do in particular situations (Cialdini et al. 1990). Previous research showed that SNs are a powerful tool to encourage prosocial behavior in various domains (Cialdini et al. 1990; Cialdini 2003; Goldstein et al. 2008). Basic UASs systems, for example, could be

integrated into social media platforms and use SNs to help users understand the importance of reporting fake news. This should motivate users to report fake news.

However, it is not clear whether basic UASs in social media are an adequate means to convey SN messages. First, the SN is communicated by a basic UAS and not by a real human. Second, the bystander-effect suggests that individuals react reservedly even when other individuals are virtually present because of a missing sense of responsibility (Latané and Darley 1970), also in the context of social media (DiFranzo et al. 2018). Third, SN messages can backfire (Cialdini 2003). On the one hand, basic UASs may apply untargeted descriptive SNs, which indicate that most users do not report fake news and, thus, demotivate users. On the other hand, SN messages may decrease news articles' believability, leading to increased reporting of real news (Schultz et al. 2007). Hence, it is a mystery whether and how far SN messages presented by a basic UAS apply to the context of fake news reporting in social media and, if they do, whether their net effect is positive or negative. IS scholars recently focused on multiple cognitive and behavioral aspects of users consuming fake news on social media. This stream of IS research suggests that subtle changes to the user interface may trigger reflection and lessen unwanted user behavior contributing to the further spread of fake news (Kim et al. 2019; Kim and Dennis 2019; Moravec et al. 2019; Moravec et al. 2020; Moravec et al. 2022). These results are encouraging for basic UAS design features such as SN messages to steer user behavior. However, none of the recent IS papers investigated SNs or other means to improve social media users' reporting of fake news. Against this background, this section focuses on the following research question: Can injunctive and descriptive SN messages provided by a basic UAS as part of a social media user interface improve social media users' fake news reporting behavior?

To answer the research question, this work derives theoretical hypotheses and conducts two online experiments. Based on a combination of theoretical arguments and empirical evidence, the results reveal that combining injunctive and descriptive SN messages as social media design features has positive effects on fake news reporting. Interestingly, descriptive SNs alone do not affect the reporting of fake news. The results contribute to research and practice. From a scholarly perspective, this work follows a recent call for more IS research on socio-technical interventions as remedies against fake news (Bernard et al. 2019) and contributes to the emerging stream of IS research on UASs. From a practical perspective, this work provides evidence for SNs as a potential design feature for basic UASs in the fight against fake news. Ultimately, using basic UASs with SNs may contribute to less fake news and, thus, lessen the societal burden of social media's adverse effects.

Hypotheses Development

A notable characteristic of social media is that decision-making is not isolated but occurs in a virtual social environment. Within such a social setting, others' attitudes and behaviors can result in SNs that strongly influence an individual's behavior (Cialdini et al. 1990; Deutsch and Gerard 1955). SNs are "an individual's beliefs about the typical and condoned behavior in a given situation" (Kormos et al. 2015, p. 480). SNs are an essential factor in the decision-making process and can lead to prosocial human behavior (Cialdini et al. 1990; Cialdini 2003; Goldstein et al. 2008). They may be communicated via SN messages visualizing or verbalizing the norm. Research distinguishes two types of SNs. First, injunctive SNs "specify what ought to be done" (Cialdini et al. 1990, p. 1015) and describe the behavior which is expected from the social environment by defining which actions are desired and which are undesired (Reno et al. 1993). Second, descriptive SNs inform about what others typically do in a particular situation (Cialdini et al. 1990). Although both SNs provide social information, they act through different motivations (Cialdini et al. 1990; Reno et al. 1993; Rimal and Real 2003). Injunctive SNs influence behavior by highlighting potential social rewards and punishments for desirable or undesirable behavior (Cialdini et al. 1991). Such moral evaluation influences the behavior even if the imaginary others are not family or friends but are general members of society (Cialdini 2007).

Although SNs mostly guide an individual's behavior, they do not always, and in every situation unfold the desired effect (Cialdini et al. 1990). The focus theory of normative conduct (Cialdini et al. 1990) assumes that an individual is more likely to behave according to a SN when the individual's attention is focused on the norm at the moment of behavior. Priming or highlighting other individuals' behavior is exemplary tools to activate a SN (Cialdini et al. 1990; Goldstein et al. 2008).

Previous research already demonstrated the importance of injunctive SNs for decision-making in different contexts such as alcohol use (LaBrie et al. 2010; Rimal and Real 2003), gambling (Neighbors et al. 2007), tax evasion (Wenzel 2004), organ donation (Park and Smith 2007), and risky sexual online behavior (Baumgartner et al. 2011). Due to these versatile influences on behavior, we suspect that injunctive SNs can also be used to improve social media users' reporting behavior. We expect that when social media users are made aware that reporting fake news is a behavior desired by the social environment, it also positively affects their decision to

report fake news. In particular, we expect that the use of an injunctive SN will lead to more reported fake news stories. Based on this, we derive the following hypothesis:

Hypothesis 3.3.1 The presence of an injunctive SN message indicating that reporting fake news is a socially desirable behavior increases the amount of reported fake news.

Contrary to injunctive SNs, descriptive SNs motivate actions by providing information about others' behavior and showing which actions are likely to be effective, adaptive, and appropriate (Cialdini et al. 1991). Others' behavior can be used as a heuristic, giving the decision-maker an information-processing advantage and a decisional shortcut in uncertain, ambiguous, or threatening situations (Jacobson et al. 2011). Previous research has demonstrated the successful application of descriptive SNs to reduce littering (Cialdini et al. 1990; Reno et al. 1993), support recycling (Nigbur et al. 2010), improve voting behavior (Gerber and Rogers 2009), promote towel reuse (Goldstein et al. 2008), increase monetary donations (Cialdini 2007; Croson et al. 2009), and to encourage healthier eating (Robinson et al. 2014). Based on these results, we suspect that the demonstration of others reporting content on social media platforms improves users' reporting behavior. Therefore, we suspect that the number of reports is increasing compared to no SNs message. This assumption leads to the following hypothesis:

Hypothesis 3.3.2 The presence of a descriptive SN message indicating that other people report fake news on social media increases the amount of reported fake news.

In addition to the isolated consideration of the two types of SNs, a combined application is also possible. In a study on recycling behavior, Cialdini (2003) demonstrates that the joint use of injunctive and descriptive SNs leads to a particularly high intention to recycle. Similarly, Schultz et al. (2008) reveal in a field experiment on towel reuse that combining the two types of SNs has a more substantial impact on behavior than just using one of them. Thus, previous research shows that although both types of norms use different motivations (Cialdini et al. 1990; Reno et al. 1993; Rimal and Real 2003), their simultaneous application can produce additional positive behavior effects. Focus theory suggests that the presence of a descriptive norm message can additionally direct the user's focus to the injunctive norm (Cialdini et al. 1990). Applying these insights to the case of fake news reporting, we expect that the presence of a descriptive SN message describing other users have reported a news post will further focus a user's attention to the injunctive SN, describing that reporting is a desired behavior, which consequently increases the number of fake news stories reported. Therefore, we expect that the

number of reports will exceed the number of reports when only applying one SN message at a time (i.e., injunctive or descriptive). Hence, we formulate the following hypothesis:

Hypothesis 3.3.3 The simultaneous presence of injunctive and descriptive SN messages indicating that reporting fake news is a socially desirable behavior and that other people are reporting fake news leads to a higher amount of reported fake news than the use of only one of the two types of SNs.

However, a particular characteristic of descriptive SNs is that a descriptive SN's strength can be varied by adjusting the number of people who show a specific behavior. Current research demonstrates a positive correlation between the strength of a descriptive SN and its influence on behavior (Demarque et al. 2015; Kormos et al. 2015). For example, Demarque et al. (2015) reveal in an online shopping environment that information about a higher number of people who bought an ecological product leads to more other customers buying ecological products. The higher the number of people who demonstrate a specific behavior, the more this behavior is perceived as the standard way of doing things and increases the willingness to use the SN as a decisional shortcut. In combination with an injunctive SN, increasing the strength of a descriptive SN further focuses the attention of people on a desirable behavior (represented by the injunctive SN) and motivates them to behave in the desired way (Schultz et al. 2007). Vice versa, a weak descriptive SN indicating that most people do not act desirably may be counterproductive (Cialdini et al. 1990).

For fake news, we suspect that the presence of a stronger descriptive SN is focusing a user's attention on a specific post and the action of reporting. Simultaneously, the descriptive SN highlights the injunctive SN and increases the likelihood that a user is reporting a specific fake news post. Therefore, we expect that the more users have reported a particular fake news story, the greater is the likelihood that other users will also report this fake news story. However, the results of Wong et al. (2016) show that the intention to report depends on the perceived benefit of the reporting action. Users report more often when they feel that reporting also helps to solve the problem (Wong et al. 2016). Therefore, we suspect that an increase in the strength of the descriptive norm could lead to the user perceiving a little additional benefit in his or her reporting action since enough others have already reported a specific fake news story to draw the social media's attention provider. This effect counteracts the positive influence of both the descriptive and injunctive SN. We suspect that from a specific strength onward, the reduction due to the low perceived additional benefit surpasses the increase due to the descriptive SN. If

this threshold is exceeded, a further increase in strength leads to a reduction in reporting probability. Based on this, we deduce the following hypothesis:

Hypothesis 3.3.4 The strength of a descriptive SN message, while an injunctive SN is also present, affects a user's likelihood to report a news story as fake. The influence is non-monotonic and follows an inverted u-shape.

Despite the desired effect of SNs (that fake news is reported), it should be noted that also undesired effects of SNs may occur. Specifically, users might report truthful news as fake. We do not hypothesize any effects here, as this is not the focus of the present article. However, we include this undesired effect on reporting behavior in our analyses to test whether the hypothesized benefit of SNs comes at a price in terms of misleading reports.

Study 1 Regarding the Effect of Social Norms

To test our theoretical hypotheses, we conducted two online experiments, Study 1 and 2. The studies are closely related, with Study 2 delving deeper on hypothesis 3.3.2.

Methodology

Participants: We recruited participants for our experiment via Facebook groups for students. Students are an exciting group as they use social media regularly, in many cases daily (Pempek et al. 2009), which makes them particularly open to frequent exposure to fake news. Using an online (rather than a lab) experiment and sampling via Facebook are sometimes discussed as methodological weaknesses. We instead consider these design choices as strengths supporting the experiment's ecological validity for our research context. We recruited participants via German Facebook posts. The experiment was implemented in German. The screenshots presented here are a translation. We motivated participation with a raffle for three vouchers.

To ensure the responses' quality, we include only data from participants who use social media at least once a week and who completed the experiment. Explicit checks were made to ensure that participants interacted with all the experiment materials to see all the content. This approach leads us to data from 320 active social media users. The sample is strongly composed of students ($n = 293$) with an average age of 23.2 years. A total of 70 percent of the participants are female ($n = 224$). In total, 267 participants state that they use social media as a source of information on current events, public affairs, and political issues at least once a week. Previous studies also highlight the importance of social media as a news source (Newman et al. 2017).

The participants rarely report posts (mean 0.12 per day, a standard deviation of 0.51). Appendix A.10 has details.

Task: We provide the participants with a self-developed news feed to manipulate and monitor the experimental environment. To increase the degree of realism and create a familiar environment for participants, we orient our development towards the best-known news feed design developed by Facebook (Newman et al. 2017). We display content in the form of posts containing a source, an image, a headline, and a lead sentence. We exclude other information that could be used by the participants as potential references to evaluate the truthfulness of a news story according to its general acceptability (social influence bias), such as the number of likes, comments, and shares. Also, every post has three standard Facebook buttons for interaction: a like button, a comment button, and a share button. We have added a report button. Such a reporting feature already exists on Facebook, but it requires several clicks to use it. We have simplified the reporting process to reduce possible difficulties and, therefore, to separate the different treatments' effects better. Figure 9 shows an exemplary implementation of a post in the experiment.

The news feed consists of 15 posts of three different types. First, the news feed includes five fake news posts containing fake news stories that spread widely on Facebook (in German-speaking countries) and are proven to be false (Schmehl and Lytvynenko 2017; Silverman 2016). We modified the fake news posts based on common fake news characteristics such as spelling mistakes, flashy formatting, or obviously altered images to make it easier to identify them. Every fake news story is assigned to a non-existent news source to reduce authenticity further. Besides the fake news, the news feed includes five real news posts. These contain no characteristics of fake news and originate from authentic news sources. In addition to the news posts, the news feed also has five neutral posts of imaginary friends to create an environment that is somewhat realistic and increases external validity. The display order of the posts is generated randomly for each participant. Appendix A.11 contains all posts.

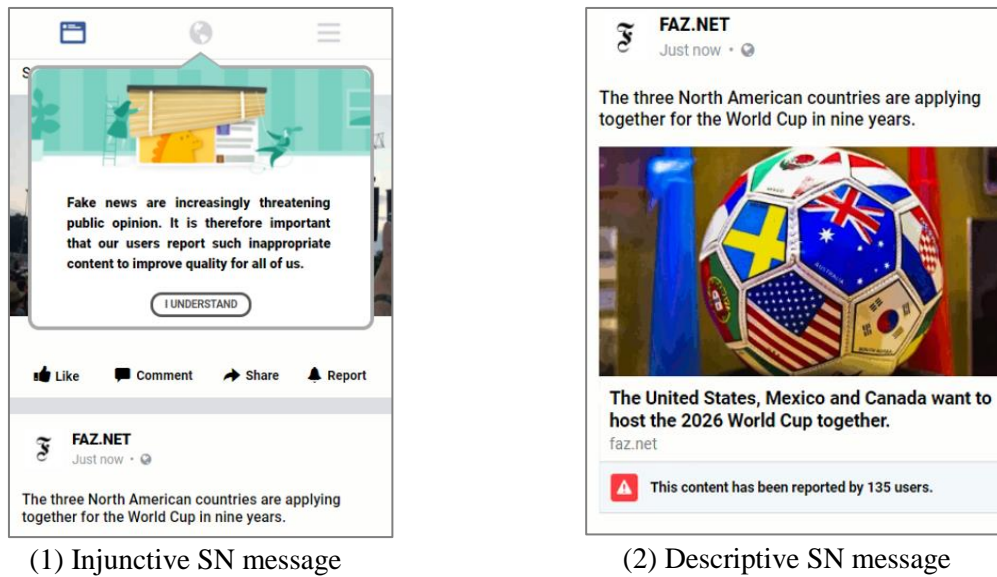


Figure 9: Implementation of the Injunctive and Descriptive SN Messages

Treatments: We confront participants with different SNs and observe fake news reporting behavior. We use a 2 (injunctive SN message: present vs. absent) x 2 (descriptive SN message: present vs. absent) between-subject design:

- (1) Control treatment: No SN messages were used in the news feed.
- (2) Injunctive SN treatment: The injunctive SN message was shown at the beginning of the news feed.
- (3) Descriptive SN treatment: Five of the ten posts with news content, whether they are fake news or real news, were randomly flagged with a descriptive SN message.
- (4) Combined treatment: Combination of (2) and (3).

To vary the descriptive SN's strength, we use a 5-level within-subject design in the treatments with descriptive SN messages. Each level occurred exactly once per participant. Based on random allocation, the number of participants per treatment ranged from 77 to 83. To test the structural equality of the four treatments, we conduct ANOVA tests (for age and usage behavior of social media) and chi-squared tests (for gender, education, and occupation) (Lee et al. 2000). The results indicate no differences at a 5 percent significance level. Thus, we assume structural equality. A detailed list of the results can be found in Appendix A.12.

Independent Variables: To be effective, SNs should be perceived as appropriate and credible (Croson et al. 2009), and they should be present in perception or the focus of attention (Cialdini et al. 1991). Based on these recommendations, we developed implementations for injunctive

and descriptive SN messages. We paid particular attention that the norm messages do not appear as an instruction by the experimenter but as part of the social media news feed. The injunctive SN message is displayed as a text at the beginning of the news feed (left-hand side of Figure 9). The text reads as follows: *“Fake news is increasingly threatening public opinion. It is, therefore, important that our users report such inappropriate content to improve quality for all of us”*. The participants have to confirm the text by clicking on the corresponding button to see the news feed. Thus, we can ensure that participants recognize the injunctive SN (the focus of attention).

We implement the descriptive SN message as additional information for a post, indicating how many users have already reported this post (right-hand side of Figure 9). By increasing the number of reports, the strength of the descriptive SN also increases. Our experiment distinguishes five distinct levels (5, 25, 125, 625, 3,125) to cover a broad spectrum without reducing credibility. The assignment of the strength of the descriptive SN to a post is random. Additionally, the number of reports for each post randomly increases by up to a maximum of 10 percent to ensure that our results are not based on a specific number of reports and that the underlying pattern and prominence of numbers are hidden from the participants.

Prior research consistently suggests a decisive role of confirmation bias in assessing social media posts (Kim et al. 2019; Kim and Dennis 2019; Moravec et al. 2019; Moravec et al. 2022), which is a prerequisite for reporting fake news. We assume that confirmation bias will also be present in our experiment. Given the broad range of topics covered by the posts in our experiment, identical posts in all treatments, and random allocation of participants to treatments, we assume identical prior beliefs and knowledge on the posts in all treatments. Thus, confirmation bias cannot be the source of treatment differences in participants' behavior.

Dependent Variables: As the primary dependent variable, we measured how much fake news a person reported. Additionally, we also measured how much real news a person reported. This approach gives us an indication of how well the fake news was recognizable as such.

Procedures: Participants were informed that the aim was to investigate the behavior of users on social media. We deliberately avoided mentioning fake news. Participants were asked how often they use social media as a news source and how much they trust this information. Then, the participants went through an interactive tutorial, during which they read a description, explicitly explaining the different features available in the news feed. The participants interacted with a single news post only used for this introduction and did not indicate a SN. The

tutorial aims to increase the participants’ awareness of the reporting feature and teach them how to report content without telling them that they should report.

Next, in the so-called natural interaction run, the participants saw the news feed with all 15 posts in random order and were asked to interact with it as if they do with their personal news feed. It is only in this phase that the four treatments differ. The natural interaction run is of the highest interest for testing our hypotheses.

Next, in the fake assessment run, the participants saw the news feed again, but all interactions were reset. The participants received explicit instruction to report fake news in the news feed and were incentivized by a higher probability of winning a voucher when performing well, specifically in the fake assessment run. The fake assessment run should not be overrated for potential effects from the multiple exposure and order of posts and the changed incentive compared to the natural interaction run. We merely use it to assess whether participants could differentiate between fake and real news if they paid close attention. As expected, the participants identified both the fake news and the real news (see Appendix A.13).

The experiment ended with brief questionnaires on social media usage and demographics and a debriefing. The debriefing informed the participants that we had manipulated different features of the posts and that these do not necessarily correspond to the real world. Figure 10 summarizes the process of the experiment. We performed a pre-test (see Appendix A.13) and report all scales in Appendix A.14.

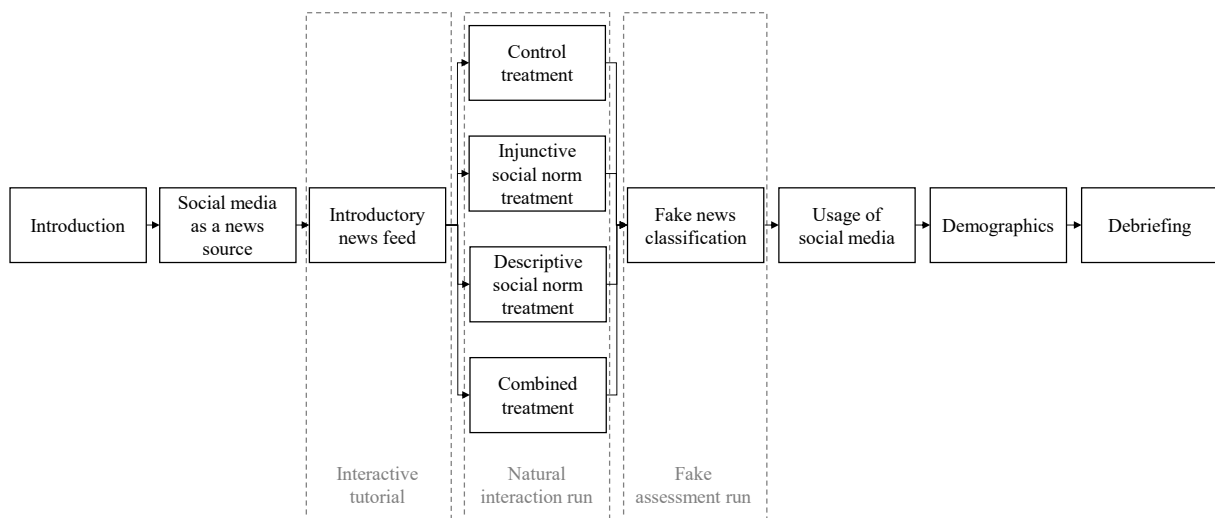


Figure 10: Summary of the Procedure of the Experiment in Study 1

Results

To measure the effect of SN messages on fake news reporting, we compare the number of reported fake news within the four treatments employing regression analysis. Due to the discrete nature of the dependent variable (integers in the range from zero to five) and non-normally distributed residuals in linear regression, we apply an ordered logistic regression (Agresti and Kateri 2011). The assumptions of an ordinal dependent variable, at least one categorical independent variable, and no multicollinearity are fulfilled by design. Additionally, we performed a Brant test and confirmed that the proportional odds assumption is satisfied (see Appendix A.15).

We use the control treatment as the baseline to analyze the differences between the control treatment and the SN treatments. We perform a rebase to use the combined treatment as a baseline to investigate the differences between the combined treatment and the injunctive SN treatment and the descriptive SN treatment. Table 5 summarizes the results.

Effect of Injunctive SN Messages on Fake News Reporting: The regression analysis shows an injunctive SN message significantly increases the probability of fake news posts being reported. Compared to the control treatment, we observe that the odds of reporting fake news significantly increase by 96 percent (i.e., it almost doubles). Transforming the odds ratio to Cohen's *d* (Sánchez-Meca et al. 2003) and applying his standard levels for small, medium, and large effects (Cohen 1969) suggests that this effect is significant but small. As we hypothesized, pointing out the importance and the desirability in the social environment of reporting fake news guides the participants' behavior toward reporting more fake news in their news feed. The results support hypothesis 3.3.1. An additional exploratory analysis of the data shows no significant effect on real news reporting while applying the injunctive SN message (see Appendix A.16).

Base Treatment		Coefficient	Standard Error	p-value	Odds Ratio (Effect Size)	Related Hypothesis	
Control	Treatments	Injunctive SN	0.674	0.326	0.039*	1.962 (small)	H 3.3.1
		Descriptive SN	0.456	0.335	0.173	1.578 (small)	H 3.3.2
		Combined	1.373	0.321	< 0.001***	3.947 (medium)	H 3.3.3
	Intercepts	0 1	0.927	0.249	< 0.001***		
		1 2	1.793	0.263	< 0.001***		
		2 3	2.312	0.276	< 0.001***		
		3 4	3.406	0.328	< 0.001***		
		4 5	4.558	0.451	< 0.001***		
	Nagelkerke's R ² : 0.068						
	Combined	Treatments	Control	-1.373	0.321	< 0.001***	0.253 (medium)
Injunctive SN			-0.699	0.292	0.017*	0.497 (small)	
Descriptive SN			-0.916	0.303	0.003**	0.400 (medium)	
Intercepts		0 1	-0.446	0.209	0.033*		
		1 2	0.420	0.209	0.044*		
		2 3	0.940	0.217	< 0.001***		
		3 4	2.032	0.270	< 0.001***		
		4 5	3.185	0.409	< 0.001***		
Nagelkerke's R ² : 0.068							

Note: Results based in an ordered logistic regression. In the first model, the control treatment is the baseline, whereas in the second model, the combined treatment is the baseline.

+ $p < 0.1$, * $p < .05$, ** $p < .01$, *** $p < 0.001$

Table 5: Comparison of Fake News Reporting in Study 1

Effect of Descriptive SN Messages on Fake News Reporting: Comparing the descriptive SN treatment to the control treatment, we see a 58 percent increase in odds of reporting fake news, but this small effect is not statistically significantly different from zero. Hypothesis 3.3.2 is not supported.

The data suggest a potential effect of the descriptive SN on the (erroneous) reporting of real news as supposedly fake. The odds of reporting real news increase by 232 percent. However, this increase also fails to reach statistical significance (p-value of 8 percent). Details on the regressions for real news reporting are in Appendix A.16.

Not supporting hypothesis 3.3.2 is counterintuitive. One reason could be that the descriptive SN was not sufficiently in the focus of attention as – contrary to the injunctive SN – participants did not have to acknowledge it. This idea is supported by the fact that only 25.1 percent of all

participants who saw a descriptive SN message stated that they used the descriptive SN messages to identify fake news in the ex-post questionnaire. Another reason could be the different effects of positive and negative descriptive SNs (Cialdini 2003; Cialdini et al. 2006). The news feed contains ten news posts, five real and five fake. In Study 1, the descriptive SN messages randomly flag five out of these ten news posts as having been reported by other users irrespective of whether the news is real or fake. Participants seeing fake news being flagged might perceive this as positive SN and might report fake news themselves. Participants seeing real news being flagged might perceive this as negative SN and might either also report real news themselves or might not engage in reporting at all. As the random allocation of the descriptive SN messages results in a mixed positive and negative SN message for most participants, we cannot single out the potentially contradictory effects of positive and negative descriptive SN with the data from Study 1. In Study 2, we separate the two types of descriptive SNs and investigate these potential effects.

Effect of Combined SN Messages on Fake News Reporting: Comparing the combined treatment with the control treatment, the odds of reporting fake news increase significantly by 295 percent (i.e., it almost quadruples; medium effect; derived from the 0.253 odds ratio reported in Table 5 for the regression with the combined treatment as the base). At the same time, the odds of reporting fake news in the combined treatment are also significantly higher than in the injunctive SN treatment (101 percent increase in odds; small effect) and the descriptive SN treatment (150 percent increase in odds; medium effect). Thus, the data support hypothesis 3.3.3.

In addition, the data reveal that the number of reports for real news is also significantly higher for the combined SN messages as compared to the absence of SN messages (control treatment; see Appendix A.16). In the control treatment, three participants reported exactly one real news. In comparison, seven participants reported one real news in the combined treatment, and five participants reported two real news. For the descriptive SN only, we found an insignificant increase in real news reporting. It appears that adding the injunctive SN amplifies this effect.

As a robustness check, we conducted additional analyses to control for the participants' awareness of the reporting feature. Appendix A.17 provides detailed information.

Effect of the Strength of Descriptive SN Messages on the Likelihood of Reporting News Stories as Fake. To evaluate the impact of the strength of a descriptive SN on the participants' reporting behavior, we use a logistic regression analysis. The binary dependent variable

describes the reporting status of a post indicating whether the corresponding participant reported the post or not. The independent variable describes the strength of the descriptive SN. Since we cannot assume a direct connection between the numerical value of the strength of a descriptive SN and the influence on behavior, we interpret the descriptive SN's strength as an ordinal factor. We focus on the combined treatment since the effects of the descriptive SN are prominent. Since each of the 82 participants of this treatment has seen five fake news posts, our sample for this analysis consists of 410 fake news. We use the Huber-White procedure to account for clustering in observations that might arise from having multiple posts in a news feed (Huber 1967; White 1980). Table 6 summarizes the results. Appendix A.18 additionally varies the baseline.

The descriptive SN's strength increases the probability of fake news posts being reported (see Table 6). We observe that, compared to no descriptive SN (strength 0), the odds are higher by 106 percent at a strength of 5, 128 percent at a strength of 25, 219 percent at a strength of 125, and 269 percent at a strength of 625. All these differences to the baseline are statistically significant; the effect sizes are small to medium. We observe indications that the weakest descriptive SN also exerts the least influence. For the strongest descriptive SN (strength 3,125), the odds are lower than for any other strength (see Table 6), statistically significantly lower than for the strength of 625 (see Appendix A.18), and not statistically significantly different from the odds for the baseline of no descriptive SN (see Table 6). Thus, we observe that the probability of reporting a post increases to a strength of 625 and then drops for the strongest descriptive SN tested in our experiment. The ex-ante hypothesized explanation is that with the increasing number of reports from other users, one's own reporting action's perceived benefits are reduced (Wong et al. 2016), leading to this inverted U-shape. Thus, the results support hypothesis 3.3.4.

Nagelkerke's Pseudo R^2 for this logistic regression is low. Clearly, other factors beyond the descriptive SN affect reporting. Likely the content, the source, and other post-level characteristics influence reporting along with individual differences among participants. As each headline and news source is used only once, our data does not allow controlling for headline or source effects. Also, we examined how the strength of a descriptive SN affects the reporting of real news. No significant differences could be observed (see Appendix A.18).

In summary, data from Study 1 supports hypothesis 3.3.1, hypothesis 3.3.3, and hypothesis 3.3.4. We do not find empirical support for hypothesis 3.3.2 so far. A reason might be that for

most participants, the specific implementation of the descriptive SN mixed elements of a positive and a negative descriptive SN. Thus, it is up to Study 2 to further test hypothesis 3.3.2.

Strength	Coefficient	Standard Error	p-value	Odds Ratio (Effect Size)	Related Hypothesis
Intercept	-1.473	0.214	< 0.001***	-	H 3.3.4
5	0.723	0.313	0.021*	2.061 (small)	
25	0.823	0.389	0.035*	2.277 (small)	
125	1.160	0.329	< 0.001***	3.190 (medium)	
625	1.306	0.311	< 0.001***	3.691 (medium)	
3,125	0.570	0.353	0.106	1.768 (small)	
Nagelkerke's R ² : 0.069					

Note: Results based in an ordered logistic regression. The baseline is a strength of 0.

⁺ $p < 0.1$, * $p < .05$, ** $p < .01$, *** $p < 0.001$

Table 6: Probability of News Post Reporting Based on the Descriptive SN

Study 2 Regarding the Effect of Social Norms

Study 1 tested hypothesis 3.3.2 without differentiation between positive and negative descriptive SNs. As mentioned above, seeing fake news posts flagged by other users can have a different effect than seeing real news posts flagged. Study 2 investigates this further.

Methodology

Study 2 is an online experiment similar to Study 1, with a notable difference in the treatments.

Participants: Recruiting procedures and filters regarding complete and valid data sets were like in Study 1, only using different Facebook groups to attract other participants (see Appendix A.13 for details). The sample comprises 157 participants who use social media more than once a week. Again, the sample is strongly composed of students ($n = 116$) with an average age of 25.6 years. About 60 percent of the participants are female ($n = 94$). In total, 142 participants state that they use social media as a news source at least once a week. Furthermore, 145 do not report any fake news or report a relatively minimal amount compared to their social environment. For more details, see Appendix A.10.

Task: The task was the same as in Study 1. However, we modified the original fake news to decrease the level of authenticity further. Emoticons were added to two of the fake news, and the term “German Armed Forces Mission” was changed to “War mission.” We replaced one

fake news completely because the participants in Study 1 did not recognize it well. We replaced the real news stories as they were outdated at the time of Study 2. These measures further increased the difference between fake and real news so that the positive and negative descriptive SN is more readily experienced. See Appendix A.13 for details on the adjustments.

Treatments: We used three treatments: (1) control treatment, (2) positive descriptive SN treatment (where all five fake news but no real news was flagged with a descriptive SN message, with each strength occurring once), (3) negative descriptive SN treatment (where all five real news but no fake news was flagged with a descriptive SN message, with each strength occurring once).

Independent and Dependent Variables: Identical to Study 1.

Procedures: The procedures were mostly identical to Study 1 with two additions. First, in the fake assessment run, the participants assessed for each news post (on a 5-step Likert scale) how fake or realistic it appeared. Second, previous literature on social media users' motivation to report fake news is scarce. To gain further knowledge of it, we asked participants whether they report fake news if they see them. Based on their answer, participants were asked to rate a pre-selection of possible motivation factors or hurdles on a 5-level semantic differential and to add further factors in a free text. Appendix A.14 lists all survey scales used in Study 2.

Results

Studies 1 and 2 are comparable in terms of reporting behavior: Neither the number of reported fake news nor the number of reported real news is statistically significantly different between the control groups of both studies (Chi-square test, p-values of 0.511 and 0.272, respectively).

To analyze hypothesis 3.3.2 in more detail, we compare (as in Study 1) participants' fake news reporting behavior in the different treatments by using an ordered logistic regression. Table 7 summarizes the results. We see that using a positive descriptive SN message leads to a 36 percent increase in odds compared to no SN message, but the effect size is less than small, and the effect is not statistically significant. Simultaneously, the results show that a negative descriptive SN message hardly leads to a change in fake news reporting behavior than no SN message since the odds increase by only 1 percent, and no significant effect is observed. Overall, even in the extreme setting of Study 2, no influence of the descriptive norm on the reporting behavior of fake news can be detected, which corroborates our results from Study 1.

As in Study 1, we have conducted additional analysis to control for the participants' awareness of the reporting feature (see Appendix A.17).

Base Treatment			Coefficient	Standard Error	p-value	Odds Ratio (Effect Size)	Related Hypothesis
Control	Treatments	Positive descriptive SN	0.282	0.368	0.443	1.326 (less than small)	H 3.3.2
		Negative descriptive SN	0.010	0.389	0.979	1.010 (less than small)	
	Intercepts	0 1	0.324	0.277	0.242	Nagelkerke's R ² : 0.005	
		1 2	1.024	0.288	< 0.001 ***		
		2 3	1.641	0.307	< 0.001 ***		
		3 4	1.981	0.325	< 0.001 ***		
		4 5	2.603	0.376	< 0.001 ***		

Note: Results based in an ordered logistic regression. The control group is the baseline.

+ $p < 0.1$, * $p < .05$, ** $p < .01$, *** $p < 0.001$

Table 7: Comparison of Fake News Reporting in Study 2

Additionally, Study 2 provides insights into motivation factors and hurdles of reporting fake news. Overall, 74 participants stated they would report fake news if they see them, and 83 would not. In summary, the motivation factors and hurdles to reporting are mainly intrinsically motivated and altruistic and not because of possible financial incentives. The answers from the free text have a considerable overlap with the preselected factors. However, two further hurdles could be identified. Table 8 shows the ranked results.

Motivation factors and hurdles	Mean	Median	Standard Deviation
Preselected motivation factors to report fake news			
I would like to avoid negative consequences that result from the dissemination of fake news	4.662	5	0.556
It is important to me to help other people to form their opinion based on true facts	4.432	5	0.812
A correct news landscape is important to me	4.338	5	0.848
I would like to improve the living conditions for myself and other people	3.865	4	0.984
I do not want that the relevance of the topic will be reduced by the dissemination of fake news	3.797	4	1.085
Fake news causes an unpleasant feeling in me	3.757	4	1.083
I would like to improve the quality of the social media platform I use	3.527	4	1.230
I want to share my knowledge with others	2.973	3	1.085
I hope for appreciation from my social environment	1.811	1.5	0.989
I expect a material or financial incentive	1.716	1	1.040
Preselected hurdles to reporting fake news			
I do not believe that the act of reporting has any effect or counteracts the dissemination of fake news	3.807	4	1.109
I am only a consumer of news and do not actively participate in the public discourse	3.723	4	1.130
The procedure of reporting is too complex for me	3.157	3	1.366
Reporting fake news offers me no material or financial incentive	2.783	3	1.440
I am not interested in public opinion	2.687	3	1.278
Fake news serves for entertainment	2.350	2	1.204
I've never seen fake news	2.169	2	1.228
Fake news is subject to the freedom of opinion	2.120	2	1.005
I don't know how to report fake news	2.000	1	1.269
Further hurdles to report fake news (answers from free text question)			
Lack of responsibility			
Own uncertainty			

Note: Motivation factors (n = 74) and hurdles (n = 83) measured on a 5-level scale where 1 corresponds to "strongly disagree" and 5 to "strongly agree."

Table 8: Motivation Factors and Hurdles to Report Fake News

Discussion

Fake news is a greater threat than ever before, especially with the widespread and fast dissemination on social media. We aim to contribute to the emerging body of IS research on fake news by studying socio-technical interventions as remedies for fake news (Bernard et al. 2019). Specifically, we theoretically and experimentally analyze the effect of injunctive and descriptive SNs on social media users' reporting of fake news.

Our work leverages knowledge on fake news in social media and knowledge originating from social psychology on SNs and the focus theory of normative conduct. Our results suggest that SN messages impact the reporting behavior of users in online environments. Social media users can be guided in their behavior by highlighting desirable behavior and making transparent what other users are doing. However, caution is needed, as SNs can also give rise to the undesirable behavior of reporting more real news as “by-catch.” Nevertheless, as the reporting levels of real news are substantially lower than the reporting levels of fake news, we argue that the overall effect of improving reporting by a combination of different SN messages is beneficial.

Social psychologists have already demonstrated the effectiveness of SNs. In part, our results are comparable to prior studies in different contexts regarding the effectiveness of injunctive SNs (Baumgartner et al. 2011; LaBrie et al. 2010; Neighbors et al. 2007; Park and Smith 2007; Rimal and Real 2003; Wenzel 2004). Apart from that, our results are, in part, also contrary to the findings of other researchers that descriptive SNs are an appropriate tool to motivate desirable behavior by describing the behavior of others (Cialdini et al. 1991).

The combined application of both injunctive and descriptive SN messages has already led to previous studies’ most substantial effects as it affects different motivations (Bernedo et al. 2014; Cialdini 2003, 2007; Schultz et al. 2008). We newly demonstrate that this is also the case in the domain of reporting fake news in social media and how SN messages can be integrated into social media interfaces. Following the focus theory of normative conduct (Cialdini et al. 1990), descriptive SN messages seem to focus users’ attention on the injunctive SN (when this is sufficiently present) and thus improve users’ reporting of fake news posts.

Also, the combined application of injunctive and descriptive SNs is more practical than just applying injunctive SNs. Our injunctive norm message was displayed when the news feed was loaded and then clicked away by the user. This procedure is not very user-friendly; therefore, a social media provider would likely not show the message every time the platform is accessed. However, the descriptive SN message continually reminds the injunctive SN’s user, thus redirecting the user’s focus back to what is “ought to be done” (Cialdini et al. 1990).

Our results provide insights that indicate a threshold exists regarding the positive effects of a descriptive SN. As Wong et al. (2016) already showed, the intention to report depends on the reporting action’s perceived benefit. This finding does not fit with other studies that show a positive correlation of behavior and strength of a descriptive SN (Demarque et al. 2015; Kormos et al. 2015). However, these studies examined effects in scenarios such as online shopping

(Demarque et al. 2015), which promises a different individual benefit than reporting fake news on social media.

The general implication of our results for research is that researchers working on fake news should consider the effects of SNs. Researchers engaging in behavioral research should integrate SNs in their theorizing and examine the prevalent SNs in the populations they study to understand behavior better. Researchers engaging in design science research should be aware of the power of SNs to steer user behavior. They may leverage this power in their socio-technical designs and should be cautious of the side effects of invoking SNs unintentionally. Furthermore, to rigorously evaluate their designs, researchers should keep in mind that SNs can be an explanatory variable of user behavior in social media environments and, thus, might measure SNs to control for their effect. Finally, it is essential to remember that publishing about phenomena like reporting behavior in social media also signals a descriptive SN and may affect user behavior.

Contribution to Theory

The results have multiple implications for research. First, we provide theoretical and empirical support for SN messages' role as an effective approach to guiding social media users' reporting decisions toward reporting fake news, but SN messages are not always effective. On the one hand, injunctive SN messages are a motivational tool highlighting moral evaluation in social media environments and affecting user reporting behavior. On the other hand, empirically, we do not find a significant effect when we apply solely descriptive SN messages in social media environments. The combination of both types of SN messages is most effective.

Second, we provide further evidence that descriptive SNs can also backfire. Users in social media environments decrease in reporting fake news when a descriptive SN message's strength exceeds a threshold. The explanation may be that users lose their motivation to report fake news due to the reduced benefit, as many users have already reported the message.

Third, we find no empirical evidence that SN messages have a negative effect on the reporting of real news. Although our data do not show a significant effect, we cannot exclude that with increasing strength of a misplaced descriptive SN message, real news becomes more likely to be reported as fake news.

Recent IS research shows that social media interface design affects users' perception of and engagement with fake news (Kim et al. 2019; Kim and Dennis 2019; Moravec et al. 2019;

Moravec et al. 2022). Moravec et al. (2022) showed that asking users to rate articles' truthfulness pushed them to think more critically about these articles' honesty and other articles. In their study, the evaluation question was part of the experimental design and not intended as a potential permanent design element of a social media interface. Furthermore, flagging fake news as "disputed by 3rd party fact-checkers" (Moravec et al. 2019, p. 1345) can increase users' cognitive activity – but the effectiveness depends on how the information is presented to the user (Moravec et al. 2019; Moravec et al. 2020). Furthermore, prior IS research suggests that providing different variants of source ratings positively affects users' cognitive and behavioral processes concerning fake news (Kim et al. 2019; Kim and Dennis 2019; Moravec et al. 2022). Our research on SNs to fight fake news is important as it is complementary to this prior IS research. We study how to make users evaluate the truthfulness with subtle changes to the social media interface, while Moravec et al. (2022) examine this evaluation's effect once the user engages in it.

Furthermore, professional 3rd party fact-checking or even stronger measures like suppressing the occurrence of articles in users' news feeds or even deleting fake news from social media platforms complement the mechanisms of users reporting potential fakes to be flagged, suppressed, or deleted. While user reporting may be used for initially detecting fake news posts in the vast amount of social media content, subsequent reliance on expert knowledge to inspect suspicious posts may be important as truly judging a news story may require personal knowledge of the events in the news article.

The novel insight of our research is that SN messages as part of the social media interface design can positively affect users' fake news reporting behavior. Our focus on users reporting potential fakes adds to the emergent body of IS research on socio-technical interventions that address the consumption of fake news (Bernard et al. 2019). This novel focus is essential as it complements the portfolio of socio-technical interventions, which may lessen the challenges fake news poses for social media platform providers and society.

Practical Implications

Our results also have implications for practitioners who design social media platforms and interfaces. First, to encourage users to report fake news, messages should be displayed, highlighting the injunctive SN that this behavior is socially desired. Second, to further increase the number of reports of fake news, practitioners should consider combining injunctive and

descriptive SN messages to focus their users' attention. In this, the use of descriptive SNs can also have adverse effects. The result is a trade-off between a higher number of reports for fake news and the accuracy of reporting fake or truthful news. One has to decide according to a social media platform's specific objectives, design, users, and environmental influences. For this reason, one may react from a certain threshold and choose alternative approaches. For example, starting from a defined number of reports, the exact number could be hidden, and an article or source rating could be shown.

Limitations and Future Research

Our findings are subject to limitations. The generalizability of our empirical results is limited by our design of the experimental materials (inspired by Facebook, a single wording for the injunctive SN message, and a single way of displaying the descriptive SN message) and by our participants (young, German, educated, active social media users recruited via Facebook). Further, we cannot ensure that everyone has participated only once. However, no duplicates were identified among the 77 percent of participants in Study 1 and 76 percent of participants in Study 2 who provided an e-mail address for the voucher raffle. Additionally, in practice, social media users cannot always be asked to confirm the same injunctive SN message. A variation of the wording may help confront the user more often with injunctive SNs and promote desirable behavior. The motivational factors and hurdles for reporting fake news that we identified in Study 2 may support the development of injunctive SN messages. The composition of the posts' headlines, the perceived reputation of the posts' sources, the posts' images and subjects, and the users' familiarity with the news might all affect the identification of fake news. We did not explore these factors related to detection but solely focused on reporting behavior.

In addition to these limitations, we also see five areas for future research emerging from our results. First, future research might empirically test our hypotheses with more diverse or complementary samples in other forms of social media that use other design elements or have a different structure, such as forums or instant messengers. In this context, future work should also deal with the boundary conditions of the descriptive SN in more detail. Our results suggest that there is an inflection point beyond which descriptive SNs no longer improve user behavior. Such inflection points should be examined in further empirical studies. Going beyond artificial experimental settings to field experiments on live social media platforms would greatly

strengthen external validity. In doing so, researchers can develop a deeper understanding of the boundary conditions of the descriptive SN.

Second, future work should explore whether SNs from the immediate social environment have a more decisive influence, as prior research from IS (Kuem et al. 2020) and other domains (Baumgartner et al. 2011; Neighbors et al. 2007) suggests. Furthermore, other factors interact with the cognition and judgment of users' news consumption in social media, for example, familiarity with news sources, topics, or headlines (Kim et al. 2019; Moravec et al. 2019).

Third, it appears a promising direction to study the effect of gamification on fake news reporting, that is “[. . .] the use of game design elements in non-game contexts” (Deterding et al. 2011, p. 10) or “a process of enhancing a service with affordances for gameful experiences in order to support user's overall value creation” (Huotari and Hamari 2012, p. 19). Gamifying fake news reporting could take, for example, the form of social games (Fang et al. 2019) or badges (Bhattacharyya et al. 2020) as investigated in other contexts by recent IS research.

Fourth, while SN-based interventions are socio-technical tools to foster active user engagement, which allows for follow-up measures such as user source rating or third-party fact-checking, we suspect that descriptive SNs also improve the ability to detect fake news. Future research may investigate this effect.

Fifth, another approach to improve users' behavior in digital environments is the concept of digital nudging (Weinmann et al. 2016). IS researchers recently successfully modified the choice architecture to guide people, for example, through priming (Dennis et al. 2020). Hence, we assume that investigating the effectiveness of digital nudging approaches to fighting fake news in social media is promising.

Sixth, IS researchers should further develop descriptive and prescriptive knowledge about fake news in social media to fight this societal threat successfully. In addition to improving reporting behavior, it is vital to gain more insights into why users believe fake news, how recognition can be improved, and what countermeasures can be found to contain fake news? One possible direction could be to enhance the users' knowledge about specific topics by providing related but potentially contradicting articles below a social media post (Gimpel et al. 2020).

Conclusion

Overall, this research activity investigates the influence of SNs displayed by a basic UAS on users reporting fake news on social media. Basic UASs are suitable for delivering social norms

because they can provide the user with the necessary norm information without creating further cognitive load through additional interaction. This work theoretically derives and empirically validates that basic UASs can utilize SNs to cause desirable behavioral changes by guiding users towards reporting fake news. The potential for applying these basic UASs seems considerable, as existing social media applications do not have to be adapted at great expense but can be extended by adding simple basic UASs with SN messages. Therefore, they are a promising tool to support existing methods against fake news and reduce their dissemination and negative impact.

4. Interactive User Assistance Systems to Promote Health Behavior Change

Non-communicable diseases, such as heart diseases and type 2 diabetes, are the leading cause of premature deaths worldwide with 16 million cases yearly (World Health Organization 2020). The World Health Organization has focused its attempts to prevent non-communicable diseases by targeting individual health behaviors related to four major risk factors: smoking, nutrition, alcohol overconsumption, and physical inactivity (i.e., the so-called SNAP factors) (World Health Organization 2020). To facilitate change in people's health behavior, health professionals carry out so-called BCIs; that is, "interventions designed to affect the actions that individuals take with regard to their health" (Cutler 2004, p. 2).

Over the past two decades, IS have created novel ways to support health professionals in delivering BCIs and enable population-wide health behavior change at scale (Michie et al. 2017; Noorbergen et al. 2019). In this vein, to further increase a system's impact on user behavior, research introduced technology-mediated. It is shown that interactive UASs in particular have the potential to help people change their behavior. The more intensive contact between the user and the interactive UAS during the interaction can trigger different psychological reactions, which can have a beneficial effect on the BCI. However, it is not only important what kind of information the interactive UAS communicates, but also how the information is communicated. Anthropomorphic design features trigger users' psychological processes and facilitate the formation of a socio-technical relationship between the user and the interactive UAS (Kim and Sundar 2012a; Pfeuffer et al. 2019). Variations in social design affect users' perceptions differently and can influence their resulting behavior (Yee and Bailenson 2007).

One approach that integrates into everyday life through existing interfaces (e.g., smartphones, websites) involves using interactive UASs represented by DHRs, such as avatars (controlled by humans) and embodied agents (controlled by algorithms) (Aljaroodi et al. 2019; Noorbergen et al. 2019). Indeed, many studies have shown that both avatars (Peña et al. 2016; Song et al. 2013) and embodied agents (Bickmore et al. 2013b; Lisetti et al. 2013) can facilitate health behavior change. However, variations in DHRs' social role, dynamics, physical appearance, and other factors have been associated with different behavioral outcomes. Thus, it is important to understand how design decisions on DHRs social features, or social cues, influence whether BCIs succeed. In this regard, DHR social cues act as stimuli for future actions, similar to how people process social cues in real-world interactions. Users then use the DHR's social cues (mostly instinctively) to form a social relationship with the DHR (Feine et al. 2019; Fogg 2003).

Several literature reviews provide important syntheses and guidance on applying DHRs for health behavior change. For instance, previous reviews have established that embodied agents can be a valuable tool for electronic health (Montenegro et al. 2019) and argued that differences in how one designs embodied agents (ter Stal et al. 2020) and avatars (Clark et al. 2019) may impact participation rates and, thus, intervention outcomes. While they have shown that DHRs can successfully facilitate technology-mediated BCIs, no review has systematically reviewed the myriad existing social design features up to now. However, as technology-mediated interventions strongly rely on socio-technical relationships, it is necessary to explore the impact that design features have on user perceptions and, in turn, how such perceptions influence intervention outcomes. Thus, this chapter focuses on the following research question: What are the social design features, targeted psychological constructs, and behavior-change interventions in interactive UASs represented by a DHRs for SNAP health behavior change in healthy populations?

This chapter addresses this question by reviewing the academic literature. In particular, this work synthesizes knowledge on UASs design available from existing studies that have empirically tested the impact of specific social design features of DHRs on the outcomes of BCIs in the SNAP domain. Accounting for the targeted research's interdisciplinary nature, 10 different databases were searched. 60 relevant papers published between January 2005, and February 2021, in computer science, health, and psychology outlets were identified. With this review, this chapter summarizes in a structured manner the most widely employed social design features, the targeted constructs in user perception, the employed BCI functions, and the evoked changes in users' behaviors. This knowledge can be used to improve the design of future interactive UASs. Further, knowledge gaps and directions for future research are discussed.

4.1. Theoretical Background

Related Reviews

Existing reviews demonstrate that DHRs have attracted broad use in different application areas, such as education and health (Aljaroodi et al. 2019), and that they can play a promising role in health applications (Clark et al. 2019; Kramer et al. 2020; ter Stal et al. 2020). Such reviews include work that has summarized DHR adoption in different fields for health promotion, disease management, and clinical psychology (Laranjo et al. 2018; Montenegro et al. 2019; Provoost et al. 2017). In terms of DHR design, reviews on embodied agents (ter Stal et al. 2020)

and avatars (Clark et al. 2019) emphasize looks, visual behavior, and, for embodied agents, language output as key factors that influence behavior in the electronic health context (Clark et al. 2019; ter Stal et al. 2020). In their scoping review of agents for coaching healthy lifestyles, Kramer et al. (2020) identified the need to 1) integrate users in the design process and 2) clarify the underlying theoretical foundations and persuasive tactics in evaluation reports. However, to the best of our knowledge, no existing review has summarized the most widely employed social design features in DHRs, the targeted constructs in user perception, the employed behavior-change intervention functions, and the evoked changes in users' health behaviors.

Foundations of Health Behavior Change

BCIs refer to “coordinated sets of activities designed to change [...] behavioral patterns” (Michie et al. 2011, p. 1). Various theories for behavior change exist (Pinder et al. 2018). Stage-based models, such as the transtheoretical model of behavior change (Prochaska and Velicer 1997), suggest that behavior change happens in discrete stages that the individual needs to undergo step by step. The perception of different social variables during the interventions, such as perceived risk or personal relevance, thus needs to be tailored to the recipient's stage of change (Lippke and Ziegelmann 2008). Continuum models, such as social cognitive theory (Bandura 1986), combine linear predictors (e.g., intentions or attitudes) to assess behavior likelihood and provide possible determinants for behavior change (Lippke and Ziegelmann 2008). As a common ground, the theories describe that BCIs happen in a social context and socio-cognitive variables determine whether an intervention succeeds or fails. However, “[e]ven when one or more models or theories are chosen to guide the intervention, they do not cover the full range of possible influences” (Michie et al. 2011, p. 2). As a synthesis based on various existing theories, Michie et al. (2014) proposed the behavior change wheel as a systematic guide for BCI design. Researchers have used it extensively in the health promotion domain due to its simplicity and accessibility (Noorbergen et al. 2019).

The behavior change wheel comprises three layers: 1) the sources of behavior combined in the capability, opportunity, motivation, behavior (COM-B) model, 2) nine behavior-change intervention functions to affect the behavior, and 3) policies that enable the BCIs. We focus on the first two, as the definition of policies falls outside the scope of designing a DHR-based BCI. The COM-B model posits that the continuous interaction between an individual's capability, opportunity, and motivation generates behavior and vice-versa (see arrows in Figure 11). Capability describes the physical skills, stamina, or strength (e.g., ability to exercise), and

psychological skills, knowledge, or mental strength one needs to perform the behavior or engage in the necessary mental processes (e.g., healthy nutrition knowledge) (Michie et al. 2014). Opportunities can be physical, provided by the surroundings, such as time, triggers, resources, locations, barriers (e.g., an environment rich in healthy foods), or social in nature, such as interpersonal influences, social cues, norms (e.g., social contacts discouraging smoking) (Michie et al. 2014). Motivation can evolve from reflective processes (i.e., planning and evaluation) or happen due to automatic processes (i.e., reflexes or emotions) (Michie et al. 2014). BCIs can affect the COM-B elements. They can target one or multiple COM-B elements and build on intervention functions (i.e., coercion, education, enablement, environmental restructuring, incentivization, modeling, persuasion, restriction, and training) (Michie et al. 2011; Michie et al. 2014). Figure 11 provides a simplified illustration of the relationship between BCIs and COM-B. One needs to tailor BCIs to users for them to work effectively (Kreuter et al. 2013). Such tailoring may involve simple measures, such as personalizing communication to the user’s name (Kankanhalli et al. 2021). However, one can also use advanced measures, such as adapting feedback messages or the BCI provider’s gender (Lisetti 2009) or tailoring the game strategy in a healthy eating game to the user’s personality (Orji et al. 2017). Importantly, design aesthetics and unobtrusiveness influence persuasiveness, dialogue support, credibility, and, finally, user adoption (Lehto et al. 2012). Thus, intervention uptake depends not only on the intervention design and targeted behavior but also on the DHR’s social design, which refers to its role, looks, social behavior, and user perceptions thereof.

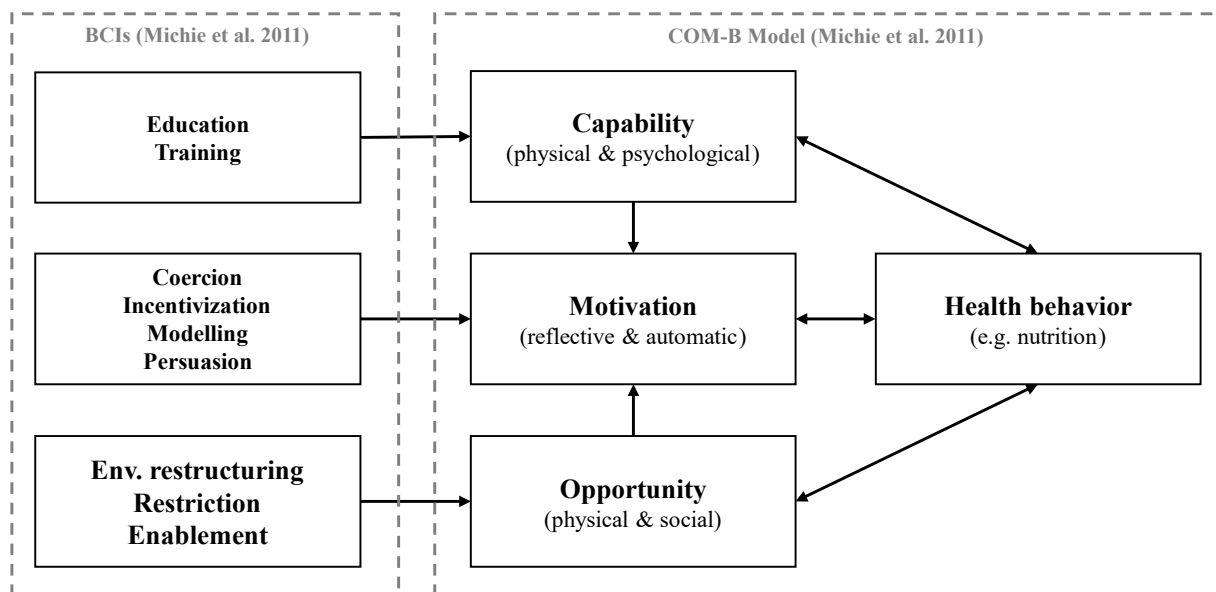


Figure 11: BCIs and COM-B Model (Adapted from Michie et al. (2011))

Social Design Foundations of Digital Human Representations

For technology-based BCIs, the user-technology relationship plays a critical role in short-term compliance and long-term adherence (Bickmore et al. 2005b; Bickmore et al. 2010). In this context, the computers-are-social-actors paradigm states that users mindlessly apply social heuristics in human-computer interaction, which triggers instinctive, automatic responses in the users, and the development of a socio-technical relationship between user and involved system (Fogg 2003). With their implicit human features, DHRs provide social cues that users know from real-world interactions and instinctively draw on to build social relationships (Feine et al. 2019; Fogg 2003). These relationships can motivate users to perform a certain behavior (Fogg 2003). For example, in their study, Cafaro et al. (2016) found that users in a greeting encounter with a digital museum guide agent quickly developed an initial impression of the agent's personality and attitude, and that this impression influenced the likelihood and frequency with which users would further use the agent. Furthermore, they found that adapting the agent to the users increased interactions with the agent (Cafaro et al. 2016). Thus, we need to understand the underlying psychological mechanisms for users to increase their readiness to interact with an agent or avatar to facilitate DHR-mediated behavior change.

Social cues refer to design features that present a salient information source and trigger social reactions in users (Feine et al. 2019; Fogg 2003). Because individuals process social cues automatically, the cues influence behavior mostly unconsciously (Fogg 2003). Fogg (2003) suggested that one can separate social cues in computing into five primary types: language, physical, psychological, social dynamics, and social roles⁴. These cues inherently form the foundation for how users perceive DHRs, such as their attractiveness, personality, and persuasiveness (Fogg 2003). For example, interaction speed and emotionality impact how users perceive an agent's empathy (Bickmore et al. 2005b; Klaassen et al. 2013a; Lisetti et al. 2013). Further, the facial similarity between a user and an agent can affect the extent to which the user perceives the agent as helpful (van Vugt et al. 2008). The Proteus effect can cause users to adapt their behavior to better match their avatar's social design (Yee and Bailenson 2007).

We identify four main components that conceptualize the relationship between DHR design and behavior change (see Figure 12). First, the availability of and interplay between social cues

⁴ One can distinguish several self-concepts (Higgins, 1987). Self-avatars can function as 1) the actual-self (i.e., as representing the user's actual personal attributes), 2) the ideal-self (i.e., as representing the user's hopes and aspirations), 3) the future-self (i.e., as modeling the user's future state if they continue a certain behavior), or 4) the ought-self (i.e., as representing the expectations or responsibilities other people impose on the user).

form a DHR’s social design. Second, via changes in user perception (e.g., perceived similarity), a DHR’s social design affects the social relationship between the user and the DHR (Feine et al. 2019). Third, changes in user perception may support or hinder the extent to which a BCI can effectively change a user’s capability, opportunity, motivation, and, finally, behavior (Michie et al. 2011). Next, we build on these four components to organize the results from our structured literature review. Specifically, we elaborate on different design variations, their effects on user psychology, and the resulting influences on BCIs and behavioral outcomes.

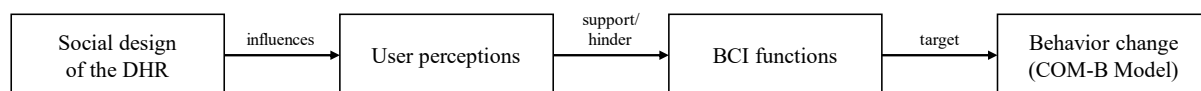


Figure 12: Structure Underlying the Present Review

4.2. Methodology

In line with our overarching research question, we focus on studies that have evaluated DHRs empirically with human users and focused on facilitating health behavior change in the SNAP domain. Following Kitchenham and Charters (2007), we divided the review process into the stages of plan, conduct and review (see Figure 13).

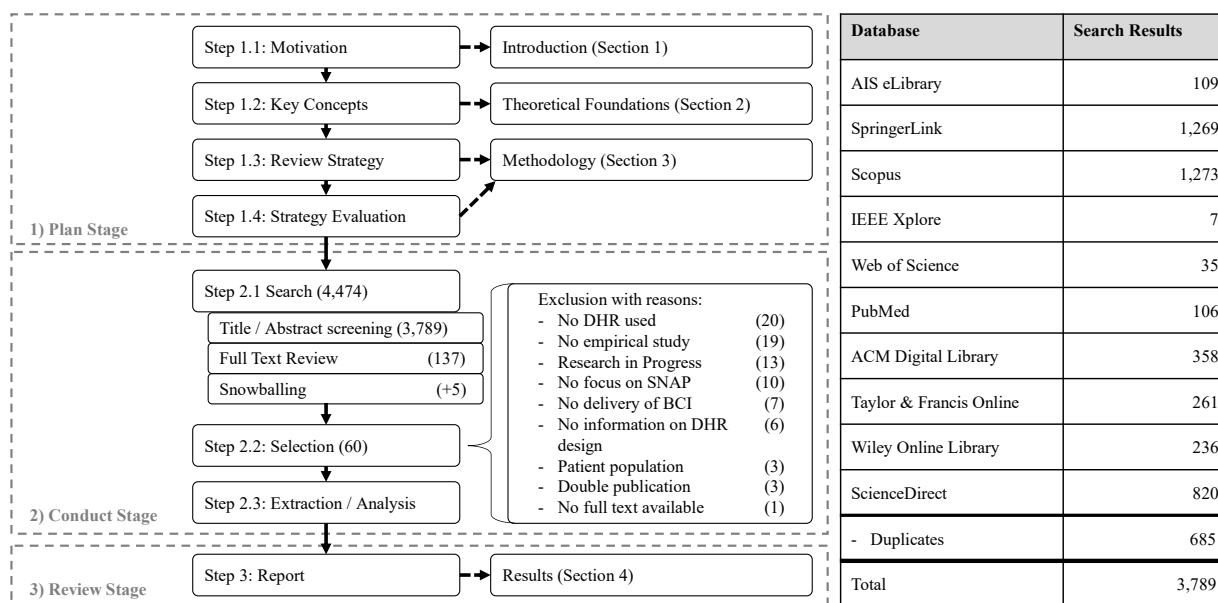


Figure 13: Stages of the Structured Literature Review

Study Selection Criteria

We include papers that 1) used at least one DHR of the self, a digital doctor, counselor, coach, friend, or similar human role (i.e., excluding animals, plants, phantasy figures, and physical

robots), 2) focused on SNAP behavior change for healthy populations (excluding papers that focused on patient populations with special requirements often connected to chronic diseases such as diabetes, Parkinson's disease, or HIV; despite being a chronic condition, we include obesity due to its close connection to nutrition and physical activity), and 3) tested the DHR in an empirical setting with humans (excluding papers that measured behavioral patterns but did not focus on changing behavior and/or did not empirically evaluate the DHR in human populations). We included all peer-reviewed journals and full-text conference publications written in English.

Search Strategy

We conducted an initial search on Google Scholar to explore the field using the search query “(avatar OR ‘embodied agent’) AND ‘behavior change’ AND health”. We reviewed the obtained results and noted relevant terms to develop our search term. We extended the search string with terms describing the role that DHRs took, such as virtual/digital advisor or virtual/digital coach. Furthermore, we concretized the search string on SNAP factors by including terms often used in concordance with the single behavioral patterns associated with SNAP. After selecting 10 well-established databases for literature in the IS and medical context (i.e., AIS eLibrary, SpringerLink, Scopus, IEEE Xplore, Web of Science, PubMed, ACM Digital Library, Taylor & Francis Online, Wiley Online Library, and ScienceDirect), we performed a full-text search in February 2021⁵. We used search alerts to stay updated on new findings from the databases.

The full-text search resulted in 3,789 unique papers in total. We screened titles, abstracts, and keywords against the defined study selection criteria (agreement rate: 94.62%). They resolved discrepancies via discussion with the first author (204 papers: 79 included, 125 excluded). We reviewed the remaining 137 papers in full and added five additional papers via the snowballing

⁵ We used the following completed search string: “(avatar OR ‘embodied agent’ OR ‘embodied conversational agent’ OR ‘mirrored self’ OR ‘virtual relational agent’ OR ‘digital adviser’ OR ‘digital advisor’ OR ‘digital coach’ OR ‘digital health coach’ OR ‘digital human representation’ OR ‘digital self representation’ OR ‘virtual adviser’ OR ‘virtual advisor’ OR ‘virtual coach’ OR ‘virtual health coach’ OR ‘virtual human representation’ OR ‘virtual self representation’) AND (‘behavior change’ OR ‘behaviour change’ OR ‘change in behavior’ OR ‘change in behaviour’ OR ‘behavioral change’ OR ‘behavioural change’ OR ‘lifestyle change’) AND (smok* OR cigarette OR tobacco OR nutrition OR eat* OR food OR diet* OR fruit OR vegetable OR alcohol OR drink* OR ‘physical activity’ OR ‘physical inactivity’ OR sport OR exercise OR walk* OR obesity OR obese OR ‘weight loss’ OR overweight OR ‘healthy lifestyle’ OR sedentary OR ‘health promotion’ OR exergam*)”. Due to length restrictions for the search string on ScienceDirect, we had to split the search string into multiple parts and conduct multiple searches. We then filtered duplicates that occurred during searches on ScienceDirect before joining the search results with results from other databases.

method. In the full-text review, we excluded 82 papers based on the reasons shown in Figure 13. We then analyzed the findings of the final corpus of 60 papers along with the components of DHR design, psychological constructs, intervention types, and behavior change (see Figure 12). Appendix B.1 provides an overview of our results. To evaluate and ensure the results' quality, we conducted a risk of bias analysis (see Appendix B.2).

4.3. Results

In this section, we present the results from our review. We describe DHR's social design variations, which we cluster according to Fogg's (2003) five principle social cue types. We found that the social design variations have a multi-faceted influence on user perceptions and intervention uptake. We report the user perceptions researchers investigated with intervention functions (Michie et al. 2014). Finally, we summarize the behavioral outcomes resulting from the BCIs.

Results of Social Design of the DHR

Social Role

Most papers (31 papers, 51.7%) used embodied agents closely followed by avatars (25 papers, 41.7%). Only four studies (6.7%) combined both DHR types. For example, users meet with agents via self-avatars in a health game (Kim and Sundar 2012b; Thomas et al. 2015). We consider these papers as belonging to the avatar papers henceforth.

All 29 avatar papers employed self-avatars. Six papers additionally compared the influence of other-avatars as opposed to self-avatars explicitly (Ahn et al. 2014a; Ahn 2015; Fox and Bailenson 2009; Navarro et al. 2020b; Peña et al. 2016; Peña and Kim 2014). One study used self-avatars and other-avatars for users to interact with a real-life nutrition or fitness professional for one hour per week (Johnston et al. 2012). Four studies found that tailoring health messages with self-avatars, as compared to other-avatars, increases BCI effectiveness (Ahn et al. 2014a; Ahn 2015; Fox and Bailenson 2009; Navarro et al. 2020b). They found both the self-avatar's and other-avatar's body size to influence physical activity. Avatar users were most active when both avatars looked physically fit, while showing an obese opponent to a user with a normal self-avatar resulted in the least activity (Peña et al. 2016; Peña and Kim 2014).

Various self-concepts emerged in self-avatars design. While two studies focused on actual-self (Napolitano et al. 2013; Thompson et al. 2016), three studies compared future- and actual-self

(Fuchs et al. 2019; Schmeil and Suggs 2014; Song et al. 2013), five compared actual- and ideal-self (Jin 2009; Johnston et al. 2012; Kim and Sundar 2012b; Lyles et al. 2017; Navarro et al. 2020a), and two compared ought-/future-, actual-, and ideal-self (Koulouris et al. 2020; Sah et al. 2017). However, most self-avatar studies did not specify the self-concept (17 of 29 / 58.6%). Sah et al. (2017) found that the ought-self promoted health consciousness more strongly than the ideal-self and actual-self. At the same time, the ideal-self yielded higher immersion than the actual-self (Jin 2009). The future-self-avatar functions as a personal model that illustrates the future consequences of a current behavior. It positively impacted interventions when projecting negative (Song et al. 2013) and positive (Schmeil and Suggs 2014) future consequences. Multiple studies took different approaches towards who created the user's avatar: the user (e.g., Jin 2009; Lyles et al. 2017) or the research teams with no modification from the user (e.g., Fox et al. 2009; Fox and Bailenson 2009; Napolitano et al. 2013). However, no study in our review corpus explicitly investigated the effect that avatar creation may have on bonding. However, findings in the studies that compared actual-self-avatars to ideal-self-avatars pointed in the direction that providing the user with the opportunity to create their own avatar is beneficial (e.g., Sah et al. 2017).

We found an even more diverse range of social roles for embodied agents. They can take the role of a friend (Bickmore et al. 2005b) who accompanies the behavior change by socially supporting positive behavior and helping when the user has a bad mood. They can act as a coach or a counselor (Abdullah et al. 2018) who provides recommendations and advice on how to change health behavior. They can function as a health professional, such as a doctor (Klaassen et al. 2013a), who provides personal health assessment. They can also act as someone who opposes health behavior change (Thomas et al. 2015), such as in a social eating practice situation where the agent tries to entice the user to eat unhealthy foods.

Further, these roles have important overlaps. For example, a counselor can also try to form a friendly relationship with the user by focusing on “hard facts” and showing interest in the user's private life and empathy (Lisetti et al. 2013). Most agent studies (28 of 31 agent studies / 90.3%) employed counselor or coach agents. However, at the same time, many DHRs try to build a social relationship with the user, such as by engaging in social talk (Bickmore et al. 2005a). Thus, other social design aspects, such as language cues (e.g., medical terminology, social talk) and physical cues (e.g., medical clothing, stethoscope), also shape a DHR's social role.

For avatars, a DHR's gender mostly matched the user's gender. For agents, most studies employed a female role (25 of 31 / 80.6%). Five agents (16.1%) matched the user's gender, and one was male (3.2%). Creed and Beale (2012) argued that users perceive female agents more favorably due to increased perceived attractiveness. Joo and Kim (2017, p. 459) conjectured that users "would respond more sensitively toward an obese female avatar than an obese male avatar". However, some researchers observed that matching agent gender (e.g., male agents for male users) yielded higher persuasiveness, although the effect was more pronounced for female users (Guadagno et al. 2007). Most agents are adults of medium age (i.e., not seniors), while studies adapted self-avatars to the user's age. Notably, no study discussed the DHR's age (e.g., gray/thin hair, wrinkles).

Social Dynamics

Following social dynamics and "knowing" unwritten patterns of interpersonal interaction can support the social user-artifact bond (Fogg 2003). Greetings and social questions about the user's feelings when an interaction begins (e.g., Bickmore et al. 2005a; Gardiner et al. 2017) and adapting content and coloring to culture-specific features (Zhou et al. 2017) represent possibilities to adapt a DHR's social dynamics. Studies reported cultural adaptation for Hispanic (e.g., King et al. 2013), African-American (e.g., Bickmore et al. 2005a), Indian (Murali et al. 2020), and Chinese populations (Zhou et al. 2017) and formulated general guidelines for Arabic DHRs (Aljaroodi et al. 2020). Nevertheless, when using culturally adapted DHRs, designers need to carefully weigh cultural cues against other factors. As Zhou et al. (2017, p. 94) stated: "[S]ince regular exercise, the topic discussed during the interaction, is not traditionally a popular theme in Chinese culture, young Chinese adults moving to the U.S. may perceive the American character as more knowledgeable, and more authoritative in the field of exercise coaching, and thus would be more willing to follow the advice offered by the American figure". Similarly, Murali et al. (2020) found that cultural tailoring works most successfully when appearance (i.e., physical cues) concurs with a culturally adopted argumentation as reflected in language and psychological cues. Taken as a whole, these findings emphasize the interrelatedness between social cues.

For embodied agents, system designers need to decide who initiates the interaction and who controls its flow (i.e., turn-taking; user or agent). In most studies, users began the conversation according to the study protocol. Other options include an acoustic signal (Bickmore et al. 2007) or a textual countdown (Fuchs et al. 2019). The agent begins by greeting the user and waits for

user input to prepare the response. During the interaction, the agent may provide hints when waiting for user replies, such as by using winks towards possible answers (Creed and Beale 2012). No study we reviewed explicitly investigated the potential effects that turn-taking has on BCI effectiveness. However, different turn-taking behaviors correspond with different perceptions of authority or dominance (Beňuš et al. 2011) and, thus, have the power to influence user perceptions and behaviors.

We also observed different overall embodiment levels, that is, body parts that the user can see (Aljaroodi et al. 2019). With only two exceptions, one face-only (Peng 2009) and one upper-body avatar (Andrade et al. 2016), the avatar studies in our review all used full-body embodiments. For embodied agents, we found a more diverse picture: 18 upper-body (e.g., Bickmore et al. 2005a; Gardiner et al. 2017; Olafsson et al. 2019), seven face-only (e.g., Creed and Beale 2012; de Rosis et al. 2006; Vainio et al. 2014), and six full-body agents (e.g., Oyibo et al. 2018; van Vugt et al. 2006, 2009). It is conceivable that the embodiment level may help DHRs form social relationships with users by influencing the perceived closeness and intimacy with the DHR if applied correctly and in correspondence to users' social expectations. However, no paper we reviewed systematically investigated the influence that embodiment level has on the social relationship to the agent or the intervention outcome.

Physical

Overall, 36 papers (60%) employed three-dimensional (3D) DHRs, while 24 (40%) relied on two-dimensional (2D) visualizations. In particular, the avatar studies primarily used 3D DHRs (26 of 29 avatar papers, 89.7%). Studies further varied in photorealism, that is, the DHR's similarity to a photographic image of a human. The reviewed studies covered the full range of this continuum from simple comic-like bodies (van Vugt et al. 2006) to comic-like DHRs personalized with photos of a user's face (Song et al. 2013), 3D photorealistic agents (Zhou et al. 2017), and self-avatars based on 3D body scans (Lyles et al. 2017).

Furthermore, 50 studies (83.3%) used some form of DHR dynamics (i.e., fluently changing their position, facial expression, or other dynamic animation). Interestingly, multiple studies investigated the effect that DHR movement had on users. They provided evidence that a DHR performing a sportive activity can motivate users to exercise better than a loitering DHR (Fox and Bailenson 2009; Joo and Kim 2017; Morie et al. 2013; Schmeil and Suggs 2014).

For self-avatars, body shape often aligns with the employed self-concept with users commonly associating their ideal-self and, hence, their self-avatar as being slenderer. Consequently, multiple studies showed that users whom a slenderer self-avatar represented would show more physical activity, motivation to exercise, and healthier nutrition behavior (Li et al. 2014; Peña et al. 2016; Peña and Kim 2014). However, as Joo and Kim (2017) noted, an avatar's body shape should align with its behavior/movements. For other-avatars, body shape exhibited a higher impact for female users than for male users. Studies identified the highest physical activity levels in situations with both a thin self-avatar and thin other-avatar (Peña et al. 2016; Peña and Kim 2014). On the contrary, larger body sizes achieved higher user preference and usage intentions (van Vugt et al. 2006, 2009).

Concerning clothing, most DHR wore workday clothes or sportswear aligned to the social role and context. Agents with a health professional's role wore medical clothing, which included a stethoscope to support the social role (Klaassen et al. 2013b; Klaassen et al. 2013a; Lisetti et al. 2015). For avatars, Navarro et al. (2020b) found users to engage in more physical exercise if their avatar wore sports rather than formal clothing. Further, some avatar studies allowed users to customize avatar clothing (Johnston et al. 2012; Kim and Sundar 2012b; Sah et al. 2017; Thompson et al. 2016; Waddell et al. 2015). Although users often request customization (Lyles et al. 2017), Waddell et al. (2015) reported that it did not positively influence physical activity levels. Controversially, previous research on avatars suggests that customization impacts identification with the self-avatar. It translates into more motivation (Behm-Morawitz 2013; Birk et al. 2016) and helps reduce attrition over time (Birk and Mandryk 2018).

Language

We first focus on the language options for user input. For embodied agents, the majority of studies (25 of 31 agent-only studies / 80.6%) implemented user input as choosing from a set of predefined answers (e.g., Bickmore et al. 2005a; King et al. 2013; King et al. 2020; Mohan et al. 2020). Olafsson et al. (2019) showed that the valence (positive/negative connotations) of the available answers influenced users' confidence in performing the targeted behavior, although users preferred to have both types of answers available. Six studies allowed free text input, although most either did not process the answers at all or did so with simple pattern matching (Bickmore et al. 2005b; Creed and Beale 2012; Friederichs et al. 2015). Only one recently published study employed natural language processing to respond to users' written input (Maher et al. 2020). In so-called Wizard-of-Oz studies (i.e., where a researcher controls the

agent), researchers allowed users to make their dialogue selection via voice input (Schulman and Bickmore 2009) or responded to textual user input by selecting dialog options (de Rosis et al. 2006). Researchers have also successfully tested system-processed voice input based on natural language processing (Yasavur et al. 2014). By contrast, most avatar-only studies (23 of 25, 92%) did not include textual user input, though some instead relied on a keyboard to control the self-avatar. Two avatar studies allowed users to communicate with each other using text input (Behm-Morawitz et al. 2016; Johnston et al. 2012).

For language output to users, studies used both written and spoken messages. For avatars, researchers used posters, menus, or other written media for verbal communication (Johnston et al. 2012; Kim and Sundar 2012b). Some studies visualized communication with other-avatars or embodied agents using text fields (e.g., Klaassen et al. 2013a; Peng 2009) or speech bubbles next to the DHR (e.g., Friederichs et al. 2015; Vainio et al. 2014; van Vugt et al. 2006). For spoken messages (typically in the user's native tongue), studies implemented a voice using either prerecorded messages (Creed and Beale 2012) or synthetic text-to-speech systems (e.g., Bickmore et al. 2005a; Lisetti et al. 2013; Yasavur et al. 2014). Various factors, such as accent, pitch, speed, and tone, require consideration. For example, the perceived politeness of agent's sound influenced short-term compliance and long-term adherence concerning the targeted behavior (Bickmore et al. 2007). Reading the messages letter by letter led to user complaints about bad glanceability (Klaassen et al. 2013b). Further, conversations with auditory output via speakers have received criticism for privacy reasons (Bickmore et al. 2007). To avoid private information being heard by bystanders, one study asked users to wear headphones during the intervention (King et al. 2013).

Psychological

Psychological cues lead users to perceive DHRs to have emotions or a personality (Fogg 2003). These cues include using social conversation in a targeted manner (e.g., asking users about their current feelings), gestures, mimics, and different verbal strategies of presentation (e.g., using humor and sarcasm). With this form of simulated social behavior, the DHR imitates behavioral patterns from human interaction and conveys emotions to users naturally (Creed and Beale 2012). As an example, agents often chat with users about their feelings and everyday experiences (e.g., Lisetti et al. 2013; Olafsson et al. 2019; Zhou et al. 2017), which can help agents form a social relationship with users (Bickmore et al. 2005b; Bickmore et al. 2010). However, repetitive dialogues and clothing can evoke negative reactions and adverse effects on

actual behaviors (Bickmore et al. 2010). Designers need to ensure that the social behavior matches the other design cues. For instance, Murali et al. (2020) found employing a culturally adapted argumentation style (e.g., collectivist or individualistic nature of the cultural background) to be more effective when the agent's physical appearance also corresponded to the same cultural background.

As a verbal strategy, humor in agents (e.g., jokes, sarcasm) yields positive user feedback (Olafsson et al. 2019; Peng 2009). Importantly, verbal strategies require the corresponding non-verbal behavior, such as smiling when telling jokes or a concerned look when talking about problems (Creed and Beale 2012; Lisetti et al. 2013). Compared to purely text-based interactions, an empathic agent showing non-verbal behavior can support intervention outcomes (Lisetti et al. 2013). Non-verbal behavior to express empathy includes facial expressions (e.g., gaze, lip movements) and gestures (e.g., pointing with the finger, shrugging the shoulders) related to the conversational content (e.g., Bickmore et al. 2005a; Creed and Beale 2012). To help a DHR mimic non-verbal behavior, designers can also use video-based expression analysis to detect the user's mental state and adapt the agent's facial expressions accordingly (Lisetti et al. 2013). For avatars, most studies used neutral facial expressions and gestures. While Kim and Sundar (2012b) allowed users to change their self-avatars' gestures, Fuchs et al. (2019) used facial expressions to visualize future consequences of health behavior.

As we have seen, studies have used social cues in various combinations and variations (social role, social dynamics, physical, language, psychological) to design DHRs in the SNAP behavior change domain. Importantly, the combination of all available cues forms a user's perception and impression, rather than only the interpretation of a single cue (Creed and Beale 2012; Fogg 2003). Hence, designers need to consider the interplay between all primary social cue types during the DHR design phase. For example, increasing the realism of an agent's physical cues will also result in higher user expectations concerning DHR's psychological cues. Further, the interplay between social cues and their relevance to the user psychology via mindful or mindless processing influences whether users form a socio-technical relationship with DHRs and whether users change their behavior (Ahn et al. 2014b; Bandura 1986; Sah et al. 2017). For messages with lower personal importance in particular, constructs such as trustworthiness, knowledgeable, or likeability become more important in evaluating the message and persuading users (Schulman and Bickmore 2009). For this reason, we now focus on the constructs that research has tested concerning user psychology and their effect on BCIs.

Results on User Perception

Table 9 summarizes the most frequently investigated constructs (i.e., at least three studies used them). Generally, we see that the studies have tested all constructs with embodied agents, but they have considered only attractiveness, enjoyment, persuasiveness, satisfaction, presence, similarity, and social distance for avatars. We can attribute this finding to the fact that most avatar studies focused on questions concerning the avatar's graphical design or social role. In contrast, agent studies included constructs predominantly related to verbal behavior in the interaction between user and agent. Furthermore, when synthesizing the results in previous studies, the interplay between and combination of constructs emerged as an important consideration for DHR design. Thus, in this section, we analyze a subset of the different psychological constructs, their interplay with each other, and the influence that DHR design has on these constructs.

Construct (closely related constructs)	Explanation	Usage
Attractiveness (aesthetics)	How much the user perceives the DHR to be visually appealing	EA: (Klaassen et al. 2013a; van Vugt et al. 2006) AV: (Jin 2009) EA+AV: (Kim and Sundar 2012b)
Credibility	How believable the DHR is to the user	EA: (de Rosis et al. 2006; King et al. 2013; Klaassen et al. 2013b; van Vugt et al. 2006, 2009) EA+AV: (Peng 2009; Thomas et al. 2015)
Ease of Use	How easy to use the user perceives the DHR to be	EA: (Abdullah et al. 2018; Bickmore et al. 2005a; Bickmore et al. 2013a; King et al. 2013; Lisetti et al. 2013; Mazzotta et al. 2009; Yasavur et al. 2014; Zhou et al. 2017) EA+AV: (Thomas et al. 2015)
Empathy (caring)	How caring and empathic the user perceives reactions from the DHR to be	EA: (Abdullah et al. 2018; Bickmore et al. 2005a; Bickmore et al. 2005b; Creed and Beale 2012; King et al. 2013; Lisetti et al. 2013; Zhou et al. 2017)
Enjoyment (entertainment)	How much the user enjoys the interaction with the DHR	EA: (Bickmore et al. 2010; Klaassen et al. 2013b; Lisetti et al. 2013) AV: (Kim et al. 2014; Koulouris et al. 2020; Li and Lwin 2016; Navarro et al. 2020a) EA+AV: (Peng 2009)
Friendliness (politeness)	How friendly the DHR appears to be	EA: (Abdullah et al. 2018; Bickmore et al. 2005a; Bickmore et al. 2007)
Knowledgeability (competence, informativeness, intelligence)	How intelligent and competent the user perceives the DHR to be	EA: (Abdullah et al. 2018; Bickmore et al. 2005a; Creed and Beale 2012; de Rosis et al. 2006; Lisetti et al. 2013; Olafsson et al. 2019, 2020; Schulman and Bickmore 2009)

Construct (closely related constructs)	Explanation	Usage
Likeability (appreciation, liking)	How much the user likes the DHR	EA: (Bickmore et al. 2005a; Bickmore et al. 2005b; Creed and Beale 2012; de Rosis et al. 2006; Friederichs et al. 2014; Lisetti et al. 2013; Olafsson et al. 2019, 2020; Yasavur et al. 2014; Zhou et al. 2017)
Naturalness (anthropomorphism, lifelikeness, realism, plausibility)	How realistic and life-like the user perceives the DHR to be	EA: (Abdullah et al. 2018; de Rosis et al. 2006; Klaassen et al. 2013b; Lisetti et al. 2013; Olafsson et al. 2019, 2020; van Vugt et al. 2006)
Persuasiveness (relevance)	How convincing of a different opinion the DHR is	EA: (de Rosis et al. 2006; Friederichs et al. 2014; Mazzotta et al. 2009; Oyibo et al. 2018; Schulman and Bickmore 2009) AV: (Ahn et al. 2014b)
Repetitiveness (habitability)	How repetitive the interaction with the DHR is	EA: (Bickmore et al. 2005a; Bickmore et al. 2010; Yasavur et al. 2014)
Satisfaction	How much the DHR fulfills the user's expectations	EA: (Abdullah et al. 2018; Bickmore et al. 2005a; Bickmore et al. 2010; Bickmore et al. 2013b; Gardiner et al. 2017; King et al. 2020; Murali et al. 2020; Olafsson et al. 2019, 2020; Watson et al. 2012; Zhou et al. 2017) AV: (Andrade et al. 2016; Napolitano et al. 2013)
(Self-)Presence (identification, representativeness)	How much the user feels correctly represented by the DHR	EA: (Lisetti et al. 2013) AV: (Ahn et al. 2014b; Behm-Morawitz et al. 2016; Fox et al. 2009; Kim et al. 2014; Koulouris et al. 2020; Li and Lwin 2016; Lyles et al. 2017; Navarro et al. 2020a; Song et al. 2013) EA+AV: (Kim and Sundar 2012b)
Similarity (resemblance)	How similar the user perceives the DHR to be compared to himself	EA: (Olafsson et al. 2019, 2020; van Vugt et al. 2006, 2009) AV: (Fox and Bailenson 2009; Morie et al. 2013; Navarro et al. 2020b; Navarro et al. 2020a; Peña et al. 2016; Thompson et al. 2016; Waddell et al. 2015)
Social Distance (personal relevance, relatedness, sociability)	How related on a personal level the user feels to be to the DHR	EA: (Bickmore et al. 2005b; Friederichs et al. 2014; King et al. 2013; Lisetti et al. 2013; Murali et al. 2020; van Vugt et al. 2006, 2009; Zhou et al. 2017) AV: (Ahn et al. 2014b; Ahn 2015)
Trustworthiness (ethics, honesty, sincerity, trust)	How much the user trusts the DHR and its messages	EA: (Bickmore et al. 2005a; Bickmore et al. 2005b; Bickmore et al. 2010; Creed and Beale 2012; de Rosis et al. 2006; Friederichs et al. 2014; Lisetti et al. 2013; Olafsson et al. 2019, 2020; Schulman and Bickmore 2009; van Vugt et al. 2006, 2009; Zhou et al. 2017)
Usefulness (helpfulness)	How much utility and practical worth the DHR has for the user	EA: (Abdullah et al. 2018; Bickmore et al. 2005b; Bickmore et al. 2013a; de Rosis et al. 2006; King et al. 2020; Lisetti et al. 2013; Mazzotta et al. 2009) EA+AV: (Thomas et al. 2015)

Note: We list constructs alphabetically. We include only constructs that at least three reviewed papers mentioned. All constructs evaluated by the reviewed studies were self-reported by the user (usually based on Likert scales). EA = embodied agent, AV = avatar, EA+AV = embodied agent and avatar(s).

Table 9: Psychological Constructs Evaluated in the 60 Papers in the Review

Likeability and Friendliness

Overall, all employed embodied agents yielded high likeability levels. An agent's ability to show emotion and empathy constitutes a key design factor for likeability. Users generally perceived agents who showed emotions or modeled user emotions as likeable (Bickmore et al. 2005b; Creed and Beale 2012; Lisetti et al. 2013). Lisetti et al. (2013) found a link between empathy and positive user perceptions. Agents can convey empathy using language cues and psychological capabilities (e.g., simulated emotion using gestures and facial expressions) that show they care for the users' situation. Caring for the users' situation goes hand-in-hand with understanding users' current feelings. Over the long run, empathy can affect the social bond between users and agent systems and, thus, also whether BCIs succeed (Bickmore et al. 2005b; Bickmore et al. 2010; Creed and Beale 2012). Interaction friendliness connects to social bond and behavior change. In an experimental setting, Bickmore et al. (2007) showed that more friendly rather than impolite interruptions achieved higher success in changing behavior over the long run. Surprisingly, cultural adaptations to the user population (e.g., young Americans with Chinese background) neither increased perceived empathy nor likeability (Zhou et al. 2017). In contrast, Indian looks and argumentation targeted to an Indian audience resulted in higher satisfaction levels (Murali et al. 2020).

Trustworthiness and Credibility

Generally, users perceived agents as trustworthy and honest based on their psychological and language cues. Agents using empathic speech and non-verbal behavior yielded higher trustworthiness (Lisetti et al. 2013), subject to the study setup's complexity and the DHR's ability to recognize the users' emotions (e.g., using face recognition) (Creed and Beale 2012; Friederichs et al. 2014). Concerning physical cues, van Vugt et al. (2006) and van Vugt et al. (2008) found that, somewhat surprisingly, users perceived more obese agents as more trustworthy, possibly due to their higher similarity and lower social distance. In general, Lee and Choi (2017) established that trust and enjoyment facilitate increased user satisfaction and intention to use. Trust closely relates to credibility, especially for psychologically complex topics (e.g., adverse health consequences). For example, an agent's facial expressions need to match the situation to support credibility adequately (Creed and Beale 2012). Similarly, Spence et al. (2013) found that emotionally correct content presentation, cultural factors, and stereotypes influence credibility.

Knowledgeability and Persuasiveness

Naturally, the information conveyed via the message content (de Rosis et al. 2006) and the interaction's structure (Bickmore et al. 2005a; Lisetti et al. 2013) influence an agent's knowledgeability. Users perceived agents as intelligent, competent, and informative about the targeted BCI topic. Also, physical features were connected to knowledgeability. For example, users perceived a more obese agent as more knowledgeable about nutrition and physical activity, which aided in persuading them about health behaviors (van Vugt et al. 2009). Overall, DHRs exhibited higher persuasiveness than pure text-based interventions (Mazzotta et al. 2009), mainly when they resembled users visually (Ahn et al. 2014b), and exhibited relational behavior (Lisetti et al. 2013). Beginning a conversation with social talk can increase persuasiveness and knowledgeability (Olafsson et al. 2019; Schulman and Bickmore 2009). Further, behavior modeling was more persuasive if designers adapted DHR gender to user gender (Oyibo et al. 2018). Lisetti et al. (2013) found a connection between knowledgeability and usefulness, which describes how users assess a DHR to enhance their everyday behavior and whether the provided messages help. Adding relational behavior significantly increased perceived usefulness (Bickmore et al. 2005b; Lisetti et al. 2013).

Presence

Research has identified the degree to which users felt as though the DHR was present in the same environment as a critical construct for behavioral outcomes (Johnston et al. 2012). For instance, a higher presence was associated with increased physical activity and healthy eating (Behm-Morawitz 2013). In our review, self-avatar studies in particular tested presence and showed that virtual faces similar to a user's actual face (Song et al. 2013), customization (Kim and Sundar 2012b), and adapting avatar physical appearance to virtual world behavior (Fox et al. 2009) increased presence and identification with the avatar⁶. Similarly, the level of "interface embodiment" (i.e., the degree to which the avatar followed real-world user movements, e.g., based on camera input) positively influenced presence, enjoyment, and participation in the BCI (Kim et al. 2014). High-presence male users copied their avatars' eating behavior in consuming more cookies, while high-presence female users reduced their cookie consumption compared to their low-presence female peers (Fox et al. 2009). Finally, studies found that presence drives

⁶ In their embodied agent study, Lisetti et al. (2013) evaluated "social presence". However, they operationalized the construct in a way closely related to the concept of naturality.

perceived personal relevance of communicated messages (Ahn et al. 2014b) and increases enjoyment and intention to use (Kim et al. 2014; Li and Lwin 2016).

Similarity and Attractiveness

Designers can achieve visual similarity to users by adapting a DHR's physical design elements. They can achieve behavioral similarity by adapting behaviors according to the target group's culture (Zhou et al. 2017) or imitating user expressions (Lisetti et al. 2013), a common strategy in health communication (Spence et al. 2013). Other means to increase similarity include adapting the DHR's physical appearance to users' size (van Vugt et al. 2006, 2009), appearance (Thompson et al. 2016), and gender (Waddell et al. 2015) and showing the desired target behavior (Fox and Bailenson 2009). Similarity decreased the social distance between users and DHRs (van Vugt et al. 2006, 2009). It drove attractiveness (Pratt et al. 2007), which was associated with higher perceived intelligence, trustworthiness, persuasiveness, and likeability (Creed and Beale 2012). Physical cues mainly determined attractiveness. For avatars, a more attractive self-avatar goes hand-in-hand with changing the social role from actual-self to ideal-self, which positively influenced behavioral outcomes (Jin 2009; Kim and Sundar 2012b). Other factors include lifelike shapes and colors adapted to user preferences.

Naturality and Social Distance

Naturality describes the degree to which users perceive a DHR as realistic and humanlike. Researchers evaluated naturality, in particular with embodied agents, and found it highly connected to physical cues (Lisetti et al. 2013). Naturality also relates to language and psychological cues. For example, while a DHR's skin color and texture should be realistic, designers should also adapt utterances to individual users (Fox et al. 2009; Friederichs et al. 2014). Naturality influenced the feeling of presence (Kim et al. 2014) and, thus, also impacted intervention uptake. More realistic DHRs, similar to the user, reduced the perceived social distance (i.e., the feeling of social relatedness to the DHR) and facilitated behavior change (Ahn 2015; van Vugt et al. 2006, 2009). Further, facial expressions, social chat, and gesture use can simultaneously decrease social distance and increase the agent's relatableness and sociability (Bickmore et al. 2005b; Lisetti et al. 2013).

Enjoyment

Researchers have evaluated various types of social design features for enjoyment. With respect to the narrative point of view, users preferred first-person stories from an agent over stories

from a third-person perspective (Bickmore et al. 2010). They perceived a low speaking speed as little enjoyable, which was connected to low intentions to use (Klaassen et al. 2013b). For psychological and physical cues, an agent showing empathy with gestures increased the perceived enjoyment compared to a non-empathic agent or a pure text-based interaction, both of which users perceived as similarly little enjoyable (Lisetti et al. 2013). Studies found a strong connection between enjoyment and both behavioral intentions and effective behavior change in digital and classic face-to-face interventions (Bickmore et al. 2010; Schneider and Cooper 2011). Further, they found that repetitiveness in agent clothing, behavior, and messages harms the enjoyment and, thus, that designers should avoid it (Bickmore et al. 2010).

Ease of Use and Satisfaction

To gain first impressions towards long-term use, system designers often evaluate users' perceived ease of use and satisfaction after they have used a system for a certain period. Overall, studies reported ease of use and user satisfaction with DHR-based interventions to be at least as high as for comparable paper-based interventions (e.g., Gardiner et al. 2017; Olafsson et al. 2019). Ease of use and satisfaction play a crucial role in system acceptance and highly depend on other constructs such as trustworthiness and repetitiveness (Bickmore et al. 2010; Kassim et al. 2012). Further, ease of use and satisfaction with a system drive usage intentions, a direct proxy for the actual DHR use and, thus, BCI uptake (Bickmore et al. 2010; Lehto et al. 2012).

Looking at the results in this subsection more broadly, we identified no single psychological construct that stands alone. The multitude of constructs available affects how users psychologically evaluate DHRs. Hence, multiple social cues in a DHR in combination cause the impact that DHR design has on user perception and cognitive evaluation. In particular, contradictions among different types of social cues (e.g., a mismatch between appearance and argumentation) may harm perception (Murali et al. 2020). An intervention's content and the psychological constructs related to the DHR delivering the BCI may support each other in helping users adopt BCIs. Hence, in the next section we look at the different intervention types that the papers in our sample used.

Results on Behavior Change Intervention Functions

The identified constructs relate to and depend on one another, which renders the way in which users perceive DHRs a multifaceted experience. In addition to a DHR's social design, user psychology and behavior are also subject to pre-intervention behavior and the provided

intervention content. Depending on the application and context, different psychological constructs may support the BCI. Yet, by diving into the most used BCIs and psychological constructs, we focus on removing some of this ambiguity.

In particular, research has shown that having a relational or empathic counselor as compared to a non-relational one positively impacted the persuasion intervention function (Bickmore et al. 2005a; Bickmore et al. 2005b; Lisetti et al. 2013; Yasavur et al. 2014). As per the COM-B model, increased persuasion leads to a direct increase in motivation with other possible impacts on behavior, capability, and opportunity via interactions among the different functions. The above sources from our sample (Bickmore et al. 2005a; Bickmore et al. 2005b; Lisetti et al. 2013; Yasavur et al. 2014) also confirmed as much. In addition, using a humorous agent may lead to increased motivation compared to a non-humorous agent (Olafsson et al. 2019, 2020; Schulman and Bickmore 2009). To improve the rate at which users take up educational content and persuasion interventions, research has increased perceived trust and credibility by including a graphically more appealing agent and social dialog (Olafsson et al. 2019; Schulman and Bickmore 2009), or by increasing the agent's physical similarity to users (van Vugt et al. 2006, 2009). Unsurprisingly, research has shown that a high satisfaction with a DHR positively influenced users' motivation and desire to continue the intervention (i.e., especially emphasizing the persuasion intervention function) (Bickmore et al. 2013b; Watson et al. 2012).

In contrast, cultural adaptation has attracted a much more controversial discourse: while associated stereotypes corresponding to an anticipated user group's physical features may also negatively affect users' motivation to change the behavior that a DHR ideally addresses, research has reported a positive effect when adapting the way in which a DHR employs argumentation. This enables users to understand the conveyed message more easily; that is, it improves users' psychological capability in the COM-B model (Murali et al. 2020; Zhou et al. 2017). To cluster the employed BCIs, we link each study to at least one of the nine intervention functions provided by the COM-B model (Michie et al. 2011). Most studies (81.7%) used multiple intervention functions; in particular, 17 papers used two, 21 papers used three, and 11 papers used four or more intervention functions. Eleven papers used only one intervention function. As the most widely employed combination of intervention functions, studies employed education, persuasion, and enablement.

Table 10 summarizes the used intervention functions grouped by the targeted COM-B components (capability, motivation, and opportunity) that Noorbergen et al. (2019) provided.

We can see that interventions using avatars mainly targeted motivation (by using the intervention functions coercion, incentivization, modeling, and persuasion) and physical capability (via training interventions). In contrast, embodied agents targeted all COM-B components. The corresponding psychological constructs that avatar studies mainly tested included presence, similarity, social distance, and (seldomly) enjoyment. Agent studies mostly used the intervention functions education, persuasion, and enablement, referring to empathy, knowledgeability, likeability, social distance, and trustworthiness in the evaluation. For training and coercion intervention functions, we identified a comparably small number of papers that investigated psychological constructs.

Construct	Capability		Motivation				Opportunity	
	EDU (28)	TRA (16)	PERS (42)	MOD (22)	COE (5)	INC (4)	ENAB (33)	ENVR (6)
Satisfaction	11 1 -	3 - -	10 2 -	- 1 -		2 - -	11 - -	2 - -
Trustworthiness	11 - -		11 - -			1 - -	10 - -	3 - -
Ease of Use	7 - 1	- - 1	8 - 1			2 - -	7 - 1	3 - -
Likeability	8 - -		10 - -			1 - -	8 - -	3 - -
Knowledgeability	7 - -		8 - -			2 - -	6 - -	4 - -
Usefulness	5 - 1	1 - 1	8 - 1			1 - -	6 - 1	2 - -
Empathy	7 - -		7 - -			2 - -	6 - -	4 - -
(Self-)Presence	1 - 1	- 3 -	1 3 1	- 7 1	- 2 -	- 1 -	1 2 -	1 - -
Social Distance	7 - -		6 - -	- 2 1	- 2 -		6 - -	1 - -
Similarity	4 - -	- 1 -	2 3 -	- 6 -	- 1 -	- 1 -	2 3 -	
Enjoyment	2 - 1	- 3 1	3 1 -	- 2 1		- 1 -	3 1 -	1 - -
Naturality	5 - -		6 - -			1 - -	5 - -	2 - -
Credibility	3 - 2	- - 2	3 - 1	- - 1			2 - 1	
Friendliness	3 - -		3 - -			1 - -	2 - -	3 - -
Persuasiveness	1 - -	1 - -	4 - -	1 1 -	- 1 -		3 - -	
Repetitiveness	3 - -		3 - -			1 - -	3 - -	1 - -
Attractiveness	1 - 1		1 1 1	- - 1			1 - -	

Note: we list constructs by decreasing number of occurrences. Numbers represent how often studies investigated a construct-intervention combination for different DHRs: # embodied agent studies / # avatar studies / # studies that used both avatars and embodied agents. Blank fields mean that no study investigated the combination. We show totals in parentheses. Single papers may investigate multiple psychological constructs and affect the numbers in multiple rows and columns. EDU = education, TRA = training, PERS = persuasion, MOD = modeling, COE = coercion, INC = incentivization, ENAB = enablement, ENVR = environmental restructuring. No study used the intervention function "restriction"; hence, we omit it from the table.

Table 10: Intervention Functions and Relationship to Psychological Constructs

Capability

Education and training interventions aim to increase users' physical and psychological capabilities (Michie et al. 2011; Noorbergen et al. 2019). We classified 28 studies (46.7%) as education interventions, that is, as increasing knowledge or understanding (Michie et al. 2014). Studies primarily used agents for this intervention function, such as to provide information

about healthy food options. Furthermore, 16 studies (26.7%) used training interventions (seven agent studies, six avatar studies, and three avatar-agent combined studies). The avatar-agent combinations used virtual worlds to practice behavior in certain situations, such as social eating at parties (Thomas et al. 2015), to impart psychological skills.

Motivation

The intervention functions persuasion, modeling, coercion, and incentivization mainly focus on increasing users' reflective and automatic motivation (Michie et al. 2011; Noorbergen et al. 2019). Most papers in our review (42 of 60 papers / 70%) employed persuasion interventions by trying to induce positive or negative feelings in users and, thereby, persuade them to behave in a healthier way. We found that 22 (36.7%) papers used modeling, which refers to providing an example to aspire to or imitate (Michie et al. 2011). Of these 22 papers, 18 used avatars, one used an embodied agent (Oyibo et al. 2018), and three used avatars and embodied agents to model behavior. Thus, we see a strong tendency towards tailoring modeling interventions by using personalized avatars. Similarly, we found that the five studies (8.3%) that used coercion interventions, which involve creating an "expectation of punishment or cost" (Michie et al. 2011, p. 7) all used avatars. We classified only four papers (6.7%) as using incentivization functions, which refers to creating an expectation of a reward. However, one could argue that many studies we classified as modeling interventions could also be seen as incentivization interventions if the user accepted the image of their future-self as an incentive.

Opportunity

Enablement, environmental restructuring, and restriction functions attempt to alter users' opportunities (Michie et al. 2011). We found that 33 papers (55%) used enablement interventions (i.e., they increased means or reduced barriers to increase the user's opportunity). We assigned primarily embodied agent studies to this category. Embodied agents enable users to have personal and professional communication about behavior change that reduces mental barriers. Further, six papers (10%) used environmental restructuring interventions that change aspects of a user's physical and/or social environment. No study used restriction interventions (i.e., setting rules to reduce the opportunity to engage in the target behavior). Instead, various studies provided educational content on self-restricting unwanted behavior or supporting self-restriction, such as setting a quit smoking date (Abdullah et al. 2018).

Results on Health Behavior Change

Targeted Health Behavior

The majority of studies targeted physical activity (29 studies / 48.3%) followed by nutrition (8 studies / 13.3%), smoking (3 studies / 5%), and alcohol overconsumption (2 studies / 3.3%). Further, 15 studies (25%) simultaneously targeted nutrition and physical activity, one study (1.7%) targeted alcohol and nutrition (Fuchs et al. 2019), and two studies (3.3%) focused on all four SNAP behaviors. However, some evidence shows that targeting multiple behaviors simultaneously may be disadvantageous: Bickmore et al. (2013a) found that their agent effectively targeted either nutrition or physical activity. However, a combined intervention targeting both behaviors saw reduced success for physical activity.

Measuring Changes in Health Behavior

Based on Palvia et al. (2015) categories, we classify the methods that the studies in our review employed to evaluate the DHRs' fidelity and measure changes in health behavior (see Table 11). Most studies used laboratory experiments (36 studies / 60%) followed by field research (23 studies / 38.3%) and surveys (7 studies / 11.7%). Note that four papers (6.7%) reported more than one method.

Behavior	Method: Measure
Physical inactivity	<p>Lab: <i>activity sensors / step count</i> (Joo and Kim 2017; Koulouris et al. 2020; Maher et al. 2020; Navarro et al. 2020b; Peña et al. 2016; Peña and Kim 2014), <i>confidence / commitment to change</i> (Thomas et al. 2015), <i>coupon choice</i> (Kim and Sundar 2012b; Waddell et al. 2015), <i>heart rate</i> (Kim et al. 2014; Navarro et al. 2020b), <i>instant rest time</i> (Bickmore et al. 2007), <i>intention to use</i> (Bickmore et al. 2007; Olafsson et al. 2020; Zhou et al. 2017), <i>intention to change</i> (Kim et al. 2014; Li et al. 2014; Li and Lwin 2016; Waddell et al. 2015), <i>psychological constructs only</i> (Schulman and Bickmore 2009; Thompson et al. 2016), <i>self-efficacy</i> (Murali et al. 2020; Peng 2009), <i>self-reported activity</i> (Fox and Bailenson 2009)</p> <p>Field: <i>intention to use</i> (Bickmore et al. 2005a; Bickmore et al. 2005b; Bickmore et al. 2010; Bickmore et al. 2013a; Gardiner et al. 2017; King et al. 2013), <i>pedometer</i> (Bickmore et al. 2005a; Bickmore et al. 2005b; Bickmore et al. 2010; Bickmore et al. 2013b; Bickmore et al. 2013a; King et al. 2013; King et al. 2020; Watson et al. 2012), <i>psychological constructs only</i> (Klaassen et al. 2013b; Klaassen et al. 2013a), <i>self-efficacy / confidence</i> (Behm-Morawitz et al. 2016; Gardiner et al. 2017; Napolitano et al. 2013), <i>self-report</i> (Behm-Morawitz et al. 2016; Friederichs et al. 2014; Friederichs et al. 2015; Gardiner et al. 2017; King et al. 2020; Maher et al. 2020; Mohan et al. 2020; Navarro et al. 2020a; Vainio et al. 2014), <i>weight loss</i> (Johnston et al. 2012; Napolitano et al. 2013), <i>vital parameters</i> (vitality score, heart rate, blood pressure, BMI) (King et al. 2020)</p>

Behavior	Method: Measure
	Survey: <i>intention to use</i> (van Vugt et al. 2006, 2009), <i>motivation / intention to change</i> (Schmeil and Suggs 2014), <i>self-report</i> (Morie et al. 2013), <i>self-efficacy, self-regulation, outcome expectations</i> (Oyibo et al. 2018)
Nutrition	Lab: <i>confidence / commitment to change</i> (Thomas et al. 2015), <i>coupon choice</i> (Kim and Sundar 2012b), <i>instant food choice</i> (Fox et al. 2009; Joo and Kim 2017; Sah et al. 2017), <i>intention to use</i> (Olafsson et al. 2020), <i>psychological constructs only</i> (de Rosis et al. 2006; Jin 2009; Mazzotta et al. 2009), <i>risk perception</i> (Ahn et al. 2014b), <i>self-efficacy</i> (Olafsson et al. 2019; Peng 2009), <i>self-report</i> (Ahn 2015) Field: <i>intention to use</i> (Bickmore et al. 2013a; Gardiner et al. 2017), <i>self-efficacy / confidence</i> (Behm-Morawitz et al. 2016; Gardiner et al. 2017; Napolitano et al. 2013), <i>self-report</i> (Behm-Morawitz et al. 2016; Bickmore et al. 2013a; Fuchs et al. 2019; Gardiner et al. 2017; Maher et al. 2020; Vainio et al. 2014), <i>weight loss</i> (Johnston et al. 2012; Napolitano et al. 2013) Survey: <i>intention to use</i> (van Vugt et al. 2006, 2009), <i>motivation / intention to change</i> (Schmeil and Suggs 2014), <i>psychological constructs only</i> (Creed and Beale 2012)
Smoking	Lab: <i>coupon choice</i> (Kim and Sundar 2012b), <i>intention to use</i> (Song et al. 2013), <i>intention / motivation to quit</i> (Abdullah et al. 2018; Andrade et al. 2016) Field: <i>intention to use, self-efficacy / confidence, self-report</i> (Gardiner et al. 2017)
Alcohol overconsumption	Lab: <i>coupon choice</i> (Kim and Sundar 2012b), <i>intention to use</i> (Lisetti et al. 2013; Yasavur et al. 2014) Field: <i>intention to use, self-efficacy / confidence</i> (Gardiner et al. 2017), <i>self-report</i> (Fuchs et al. 2019; Gardiner et al. 2017)

Table 11: Targeted SNAP Health Behaviors, Methods, and Measures

We may explain the fact that so many studies used laboratory experiments (range: 15 to 322 participants, median: 61) based on 1) the notion that some experiment setups require dedicated hardware that researchers could not provide to users for a prolonged period and 2) the higher control level in laboratory environments. Field studies (range: 6 to 4,302 participants, median: 54) provide a longer observation period and, thus, make it possible to observe changes in different COM-B model components and physical outcomes over longer periods. The duration varied from one week to one year. In particular, long-term studies reported high attrition rates over time. For example, Friederichs et al. (2015) reported an overall attrition rate of approximately 65 percent for a year. Surveys (range: 50 to 673 participants, median: 259) commonly employed interactive online questionnaires where users first had some time to interact with the DHR and answered a questionnaire afterward. The interaction with the DHR took from one to six minutes on average though some papers did not provide details.

The most widely employed outcome measure across all study types was intention to use (17 studies, 28.3%), which generally yielded high levels. This measure assumes that users can

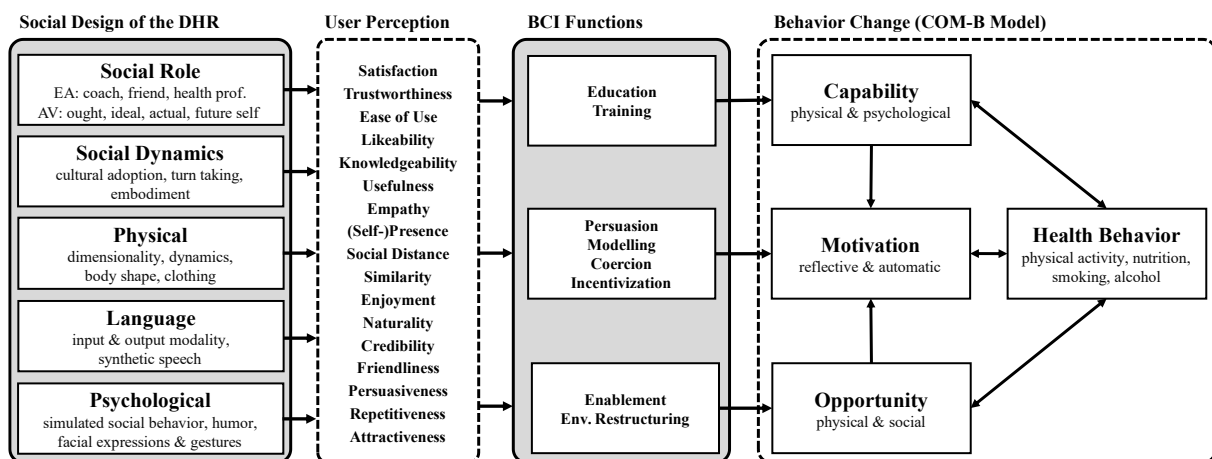
realistically project how they will use DHR for behavior change in the future. In laboratory experiments, behavioral measures often included step count/activity measures during the experiment (Joo and Kim 2017; Peña et al. 2016) and instant food/coupon choices after the experiment (e.g., Kim and Sundar 2012b; Waddell et al. 2015). Evaluations in field studies included self-reports (i.e., alcohol/cigarette consumption, food intake, and physical activity), user weight throughout the BCI, and pedometer data as quantification of physical activity. For smoking, studies also used a quit date as a behavioral measure (Abdullah et al. 2018). In online surveys, measures for health behavior included users' self-reported feelings towards behavior, such as the motivation to change the targeted behavior (Creed and Beale 2012), or outcome expectancies when continuing the DHR-based intervention (e.g., Oyibo et al. 2018).

Effectiveness of Interventions to Change Health Behavior

Overall, many studies we reviewed showed that one can effectively use DHRs for SNAP behavior change. The studies confirmed that a DHR's social design influences user perceptions and impacts an intervention's success. Compared to control groups that did not use a DHR-based intervention or used another type of intervention such as information sheets (e.g., Bickmore et al. 2005a; Gardiner et al. 2017), DHR-based intervention users reported higher fruit and vegetable consumption (Bickmore et al. 2013a; Gardiner et al. 2017), decreased self-reported food consumption (Ahn 2015), increased physical activity (Bickmore et al. 2005a; Bickmore et al. 2005b; Bickmore et al. 2010; Bickmore et al. 2013b; Bickmore et al. 2013a; Friederichs et al. 2014; Friederichs et al. 2015; King et al. 2013; Watson et al. 2012), increased exercise efficacy (Ahn et al. 2014b), and reduced alcohol consumption (Gardiner et al. 2017). Only a few studies reported that DHRs had an insignificant impact on intervention outcome compared to a control group (Andrade et al. 2016; Fuchs et al. 2019; Klaassen et al. 2013b). In contrast, other studies found DHRs as effective as other BCI deliveries such as human advisors (Johnston et al. 2012; King et al. 2020) and superior to information sheets (Gardiner et al. 2017). DHRs significantly altered participants' behavior in comparison to control groups (Bickmore et al. 2013b; Watson et al. 2012) or pre-study behavior (Maher et al. 2020; Mohan et al. 2020). DHR interventions increased confidence and motivation for physical activity and nutrition (Olafsson et al. 2019; Thomas et al. 2015). Studies reported significant differences between pre- and post-intervention user weight for avatar-based interventions targeting physical activity and nutrition (Johnston et al. 2012; Napolitano et al. 2013). Studies that compared different DHR designs and DHR behavior variations reported significant differences in behavioral

intentions and observed behaviors between the conditions tested (e.g., Bickmore et al. 2007; Joo and Kim 2017; Morie et al. 2013; Sah et al. 2017; van Vugt et al. 2006). Especially, studies reported significant differences when comparing agent-based to text-only interventions (e.g., Schulman and Bickmore 2009) and empathic to non-empathic agents (e.g., Bickmore et al. 2005a; Lisetti et al. 2013).

Overall, our results show that research has implemented different BCI types using DHRs facilitated through a range of social design features (Fogg 2003). Thereby, we can see that, depending on the DHR design, the interaction triggers various positive or negative user perceptions. For example, empathic behavior in DHRs leads to higher likeability, trustworthiness, and enjoyment during interactions with them (Lisetti et al. 2013). The triggered psychological constructs influence BCI functions' applicability and effectiveness. Likeability, trustworthiness, and enjoyment positively influence persuasion during an intervention (Bickmore et al. 2005b; Bickmore et al. 2010; Creed and Beale 2012) to motivate users to behave in a certain way. BCI functions influence the sources of human behavior (capability, motivation, opportunity) and, thereby, help individuals achieve changes in behavior (Michie et al. 2011). The results demonstrate that how one designs a DHR can have a significant impact on behavior change via the user perception constructs triggered and the BCI functions selected. Therefore, when selecting appropriate BCI features during the design phase, designers need to consider already which constructs might be beneficial or harmful in the intervention and which design features they should use accordingly. Figure 14 summarizes our findings.



Note: We show typical social design features and behavior change elements from the literature corpus. We list user perception constructs by decreasing number of occurrences. We group BCI functions by behavior change component primarily targeted; in each of three groups, we list BCI functions by decreasing number of occurrences. We do not show functions not included in the literature corpus. EA = embodied agent, AV = avatar.

Figure 14: Overview of the Results

4.4. Discussion

Summary of Findings and Interplay of Components

Over the last 15 years, researchers have created and evaluated DHR designs to facilitate technology-mediated interventions for health behavior change in the SNAP domain. Variations in the DHR's social design have yielded important differences in user perceptions that affected the interventions' efficiency in targeting users' capability, opportunity, motivation, and behavior. In this chapter, we conducted a structured literature review to establish the current body of knowledge for the role that DHR design plays in behavior change in the SNAP domain. We summarize the key relationships between DHR design, user perceptions, and intervention functions in a structured manner to facilitate behavior change (see Figure 14).

Given the myriad foci in the individual studies (e.g., physical activity, nutrition) along with the different methods (e.g., lab, field), measures (e.g., various constructs, outcome variables), and alternative BCI delivery modes (e.g., text-based, no intervention), we could not exhaustively evaluate all possible cause-and-effect relationships that the arrows indicate in the overview in Figure 14. Many possible interrelations between DHR design and user perception and between user perception and BCI interventions exist, and the extant literature has not studied all potential interrelations. Thus, we do not find it sensible to show the exact matching of design, perception, and BCI functions. Hence, this chapter aimed to summarize the different types of social cues, psychological constructs, BCI functions, and outcome variables in the SNAP health behavior change context. We conducted a risk of bias analysis of the underlying study designs (see Appendix B.2)

Based on the insights from our literature corpus and on existing frameworks, we conclude that design features (based on DHRs' social cues) cause psychological reactions from users that can support the BCI functions and behavior change. We point out initial evidence for such effects, while we emphasize context dependency of any given effect relating to DHR design, BCI, behavior change goal, and participant population. We highlight the constructs and interrelations typically studied in extant literature. Studies mainly explored improving capabilities via the effect that DHR design has on satisfaction, trustworthiness, and likeability in delivering education interventions. Studies examined these three user perceptions (satisfaction, trustworthiness, and likeability) the most regarding their effect on persuasion to increase motivation and enablement to improve opportunities. Since the studies we examined used heterogeneous designs and measures, we could not provide a numerical meta-analysis; yet,

theoretical reasoning and the cumulative empirical evidence suggest that designing embodied agents in a way that caters to these user perceptions tends to impact SNAP-related behavior positively. We lack respective evidence regarding avatars. For avatar design, studies focused particularly on the effect that (self-)presence and similarity perceptions have on modeling, persuasion, and enablement. Like before with other mechanisms for agents, theoretical reasoning and the cumulative empirical evidence suggest positive effects on behavior change. Researchers should take these findings as a starting point for understanding what has been effective so far in many studies. However, it should not limit future research on these mechanisms. The DHR's social design relates to a range of primary social cue types (Fogg 2003). So far, interventions in this area have primarily relied on self-avatars and agents and rarely examined the interplay between self-avatars and other-avatars. Self-avatars employed a range of self-concepts, such as actual-, ideal-, ought-, and future-self, whereas embodied agents acted as counselors, friends, health professionals, and behavior change opponents. Similar to the real world, the social dynamics of these interactions are vital (e.g., cultural adaptation and turn-taking).

Further, physical cues such as a DHR's body shape, clothing, dimensionality, dynamics, and photorealism affect users' perceptions of similarity, a critical construct for intervention outcomes. Researchers relied on synthetic speech, pre-recorded output, and written text for language cues. They mostly restricted user input to choosing an answer from a list with seldom free text input (e.g., as via Wizard-of-Oz studies) (Yasavur et al. 2014). In terms of psychological cues, researchers relied on humor, non-verbal behavior, and social chat that match the conversation content (e.g., gestures and facial expressions).

In our review, we identified the constructs most widely associated with DHR-based interventions. Avatar studies mainly evaluated presence and similarity. Agent studies also considered other constructs, such as attractiveness, credibility, knowledgeable, likeability, naturalness, and trustworthiness. These user perceptions influence users' intervention uptake and effectiveness through a range of psychological mechanisms. The user perceptions that a DHR's social design triggers support and/or hinder BCI functions. For instance, knowledgeable and trustworthiness facilitate education and persuasion interventions, while presence drives modeling interventions. Likely, the increased physical similarity between DHR and user and an empathic agent support the persuasion intervention function. At the same time, cultural adaptation is a controversial topic, as its effects also depend on stereotypes that designers may not consider or anticipate at the time of system design. Overall, agent studies mainly relied on

education, persuasion, and enablement, while avatar studies mainly employed intervention functions targeting motivation (i.e., coercion, incentivization, modeling, and persuasion) or physical capability (i.e., training).

Taken as a whole, the majority of the reviewed studies provided empirical support that one can successfully use DHRs for BCIs, especially compared to pure text-based. Nonetheless, readers should not understand this review as definitive advice to use DHRs for SNAP behavior change. Each time someone applies DHRs, one needs to carefully consider the unique circumstances that surround the targeted health behavior, the intervention's audience, and the possible alternatives to DHRs for delivering BCIs. Individual studies found that, for specific applications, DHRs can be similarly effective for delivering BCIs when compared to human advisors (Johnston et al. 2012; King et al. 2020) or superior to information sheets (Gardiner et al. 2017). Studies have also shown them to alter study participants' step counts significantly in comparison to non-intervention control groups (Bickmore et al. 2013b; Watson et al. 2012) or self-reported behavior in contrast to pre-study behavior (Maher et al. 2020; Mohan et al. 2020). Yet, the literature currently has neither systematically assessed the implications of the social relationship (e.g., empathy, caring) developed between user and DHR as part of the intervention nor compared the success of DHR-based BCIs to alternative delivery modes (e.g., text-based, dialogue). In particular, a comparison with the related field of human-robot interaction could benefit both fields. Robots have also become increasingly important in healthcare (Esterwood and Robert 2020) to both treat people (Olaronke et al. 2017) and deliver health interventions (Kidd and Breazeal 2008). In these cases, the social design elements influence the intervention's effectiveness. Despite different interaction modes, DHRs on a screen versus embodied physical robots, some findings we identified would also be useful for robot design (Esterwood and Robert 2020; Olaronke et al. 2017; Złotowski et al. 2015). In particular, designers should consider the relationship between social design elements, psychological constructs, and behavior change that we have elaborated on when designing human-robot interactions. Therefore, researchers should consider which constructs support the desired outcome and which social design elements would suitably promote these constructs.

Practical Implications

Based on our review, we identified various factors that system designers should consider when using DHRs for behavior change in the SNAP domain.

First, system designers need to put the targeted health behavior at the forefront of all design considerations. The DHR's social design then follows on to achieve this end. Thus, building on the structure in Figure 14, designers need to approach the subject matter "from right to left". They can do so via specifying the targeted health behavior, user group, and COM-B components (capability, opportunity, and motivation) early on. For instance, focusing on one specific health behavior rather than multiple health behaviors yields better intervention outcomes (Bickmore et al. 2013a). Accordingly, designers should choose the intervention function(s) and then prioritize the psychological constructs that support these intervention functions (Michie et al. 2014). A DHR's design should reflect these decisions to ensure the best possible intervention uptake and, finally, success in behavioral terms. Our study provides support by showing which constructs and interrelations prior work has studied as well as references to the original papers to engage with detailed knowledge regarding focal constructs and interrelations.

Second, system designers need to carefully consider the type of social relationship they intend to build between users and DHRs as it has the power to impact the intervention uptake significantly. DHRs need to build a social relationship with users and convey empathy for users to comply with an intervention over the long run. Thus, designers should design DHRs to explicitly use verbal elements (e.g., voice pitch, tone, and specific speech content) and non-verbal elements (e.g., gestures and facial expressions) that resemble human communication. This consideration of verbal and non-verbal elements in DHR design goes hand in hand with rendering the intervention as more enjoyable, the DHR as more trustworthy, and information as more credible. A promising avenue in this regard is the co-design approach as it involves end users and other stakeholders early on and in all stages of the systems design and evaluation process (Noorbergen et al. 2021).

Third, system designers need to focus on maximizing the extent to which users perceive presence and similarity to the DHR. Conveyed through physical, social dynamics, and psychological cues, presence and similarity increase an intervention's perceived relevance (Yee and Bailenson 2007) and, thus, also contribute to short-term compliance and long-term adherence. For example, avatars in interventions fostering physical activity should be shown as physically active. Also, designers can use different self-concepts to influence users' immanent self-perceptions and use them for the intervention. To enhance an avatar's similarity and attractiveness to the user, designers can implement customization to help users to bond with it and show an actual, ideal, ought, or future version of themselves.

Fourth, system designers need to carefully consider how they match intervention functions to a DHR's social design. For instance, studies building on embodied agents have primarily used education, persuasion, and enablement as intervention functions with support from the constructs, trustworthiness, satisfaction, and likeability. In contrast, studies building on avatars have so far primarily used coercion, incentivization, modeling, persuasion, and training to alter the user's motivation and physical capability. Studies have primarily used the user perception constructs (self-)presence, similarity, and enjoyment with these BCI functions, which demonstrates the need to match social design cues and psychological constructs according to the intended user perceptions that designers want to use to support the BCIs. Past research has shown that having a relational or empathic counselor as compared to a non-relational one can positively impact the intervention function of persuasion, which (according to the COM-B-model) leads to a direct increase in motivation with other possible impacts on behavior, capability, and opportunity for interactions among the different functions. Also, physical similarity plays an essential role in increasing trustworthiness and credibility, which researchers have successfully used to improve the uptake of education BCIs.

Knowledge Gaps and Directions for Future Research

Building on our review findings, we identify seven directions for future research. First, researchers need to better understand non-professional social roles for other-avatars and embodied agents. Up to now, most agents have acted as counselors or coaches, while the BCI literature suggests that non-professional roles can also support successful behavior change. This should also include the integration and contrasting of multiple roles. For example, one could combine a professional coach using an other-avatar for interpersonal communication once a week with an agent acting as an emotionally supportive friend daily. To this end, op den Akker et al. (2018) proposed the concept of a council of coaches for future research.

Further factors for future work include DHR age, gender, and gamification, which researchers have not yet systematically studied with SNAP BCIs. While many DHRs have focused on an educational and conversational approach towards communicating messages rather than for gamification, researchers have increasingly acknowledged avatars and agents as a tool for gamification. A body of research on gamification (with and without DHRs) to support health behavior change exists (e.g., Schmidt-Kraeplin et al. 2019). We see value in future research integrating DHR and gamification research to support health behavior change, especially regarding longer-term interventions where DHRs and gamification in a combined manner focus

on resolving the attrition problem that some studies in our sample reported. Gamification could, for example, support interaction with a DHR over prolonged periods required for behavior change. Conversely, DHRs could support gamification, such as being the medium for communicating feedback, stimulating competition, or increasing the emotions related to receiving badges. Besides the social role, research has neglected the influence of social dynamics during BCIs. Such social dynamics include the interpersonal distance, initiation, timing, and interaction frequency with a DHR. Even though no paper in our review studied them, it is conceivable that users may perceive a lower interpersonal distance and feel socially closer when they can see only a DHR's face or upper body as compared to a full body image.

Second, more research needs to examine the user-DHR relationship's temporal dynamics. Similar to real-world relationships, social dynamics and perceptions may change over time. As such, system designers may consider changing a DHR's social design over time, which may decrease repetitiveness, a key reason for attrition (Bickmore et al. 2010). Changing social design over time also implies the need to understand the psychological constructs in a more detailed way, which researchers can achieve by more consequently modeling user feelings. Especially, an increased use of user feedback and sensors, such as activity trackers and video cameras, may support an improved modeling of user feelings (Rouast et al. 2021). Using this additional knowledge would allow researchers to adapt a DHR according to users' preferences and perceptions and, thus, increase their interest and motivation to participate in an intervention. To this end, comparisons about the influence that different social cues have on users in different populations could further help them tailor BCIs to specific user groups, or even single users, and their current feelings.

Third, based on established literature and findings from our literature corpus, we established that how users perceive DHRs affects whether they adopt BCIs in that different psychological constructs benefit certain intervention functions. However, while we have initial evidence, broader studies and the design of behavior change applications according to the findings from our framework remain an open point for future work. We have seen a plethora of constructs, especially regarding user perception. So far, the literature does not identify clear patterns about which design features have the most importance for user perceptions, and which perceptions best support specific BCIs, and, ultimately, behavior change. We do not see a strong need to extend the list of DHR design features or user perceptions. Instead, future research should systematically explore and report the relative extent to which design features and perceptions contribute to a systematic body of design knowledge.

Fourth, we need more research on coercion and restriction interventions. As most DHR-based interventions currently focus on positive emotions, we lack research on why users adopt or do not adopt interventions that use negative emotions. For instance, we do not know whether DHRs that additionally and/or exclusively build on coercion and restriction may facilitate positive behavioral change in alcohol overconsumption and smoking. These behaviors currently lack representation as BCI targets compared to physical activity and nutrition. However, such interventions need to overcome the several challenges involved such as identifying restriction breaches, enforcing rules in the real world, and achieving acceptance among users.

Fifth, researchers should conduct more field studies to understand mid- and long-term behavior change, especially in comparison to current face-to-face interventions. At this stage, DHR evaluation primarily relies on laboratory experiments on specific design decisions. While these studies provide important knowledge to disentangle cause-and-effect relationships under controlled conditions, we need complementary studies to evaluate DHR effectiveness in the real world. The real-world context would also allow researchers to include additional stakeholders (e.g., policymakers in the planning and evaluation stages) and, thus, potentially better address larger target populations in more realistic settings. Given the focus on changing an individual's health behavior, one would need to capture the real-world BCI's interdisciplinary context. Researchers need to effectively investigate the effect DHRs may have on behavior change compared to the current standard of care, such as face-to-face interventions. While two studies have so far analyzed digital counseling with a DHR in contrast to face-to-face interventions with human advisors (Johnston et al. 2012; King et al. 2020), future studies should especially also focus on extending the knowledge in this field to provide further insights into the social relationship that emerges with users (e.g., empathy, caring).

Sixth, behavior change is not easy. Many people struggle to change their behavior even if they cognitively understand that they can and that their current behavior harms their health. Indeed, that struggle explains why researchers examine how systems can support behavior change. However, to date, we lack knowledge on the contexts, conditions, and situations that make SNAP-related behavior change particularly challenging. Having such knowledge would allow researchers to target DHRs and their intended use patterns towards systematically reducing these challenges or respecting their influence on DHRs' design and potential effect. Thus, we suggest future research to identify such challenging contexts, conditions, and situations and their interrelation with DHRs.

Finally, our review shows that future research on DHR-facilitated health behavior change should aim for more abstract (mid-range) theories. Many studies report on specific DHR designs in particular contexts. While we need such studies for the field to mature, such an overly narrow focus leads to a somewhat disparate body of literature that does not clearly show what overarching picture one can expect with a new DHR in a new context. We believe that the field would benefit from more abstract approaches that use context, BCI, target behaviors, and the like in their theorizing. For example, one could take the existence of DHR design features and BCIs as independent variables, the extent and durability of health behavior change as the dependent variable, and the health context (smoking, nutrition, etc.) as moderating variables. With this chapter, we make a step towards integrating extant knowledge regarding DHR for health behavior change. At the same time, we leave ample opportunity for primary design-oriented and empirical research to advance to more abstract theorizing.

Boundary Conditions

In our review, we examined studies that have used DHRs to deliver BCI in the SNAP domain. Hence, we explicitly excluded robots as well as avatars or agents with no human representation (e.g., virtual animals or mythical creatures). Furthermore, we focused on healthy populations to maintain comparability among the different studies without considering special needs (i.e., physical or psychological impairments related to a specific health condition). Moreover, our results pertain only to the SNAP domain and, hence, one must take care in transferring our findings to other areas. Reviewing DHRs' social design impact in other areas may also yield valuable findings for DHRs in health behavior change. We focused on covering a broad range of publications across different databases with our search string. Hence, we decided on a relatively simple search string involving various terms used across the literature. We cannot rule out that we missed a relevant publication. However, we note that we additionally engaged in snowballing, which led to five additional publications. Hence, we have confidence that our review covers many publications on the subject matter.

In particular, the COM-B model's strength lies in its broad applicability for understanding behavior change in various settings and helping one select general intervention functions (Hendriks et al. 2014; Smits et al. 2018). We built on this broad applicability to structure the different types of interventions in extant research. However, given the COM-B model's broad applicability, we note that we have not addressed all the complex relationships between capability, motivation, opportunity, and behavior. For instance, as Noorbergen et al. (2019)

have indicated, an environmental restructuring intervention may not only directly affect a person's physical opportunity to engage in healthy eating habits but also indirectly affect a person's motivation. Further, the COM-B model received criticism for not including the "crucial emotional element of wanting" as a connection between intention and behavior (Marks 2020). While one could argue that one can see "wanting" as a part of psychological skills, one should not neglect it in evaluating interventions post hoc as it helps one to understand study participants' initial situation and internalized drivers for participation. The COM-B model has also received criticism for being "too broad" to provide specific guidance for selecting intervention functions and for not indicating which policymakers one should include in developing new interventions (Hendriks et al. 2014). As a result, the COM-B model's application in practice can be time-consuming and lengthy, especially when analyzing questionnaires, and despite providing clear guidance (Ojo et al. 2019).

4.5. Conclusion

DHRs are a promising tool for the design of interactive UASs. DHRs such as avatars and embodied agents have seen increasing use for delivering BCIs that target modifiable risk factors of a person's health behavior in the SNAP domain. Based on a structured literature review, this chapter summarizes the current body of knowledge on the influence that avatars' and embodied agents' social design features have on BCIs in the SNAP domain. While the results reveal increasing evidence for general effectiveness in facilitating health behavior change overall, further research needs to better understand how the DHR of an interactive UAS compares to alternative ways to deliver BCIs. Further, existing research in this context primarily focuses on physical activity and nutrition while smoking and alcohol overconsumption have only received limited research attention. Researchers and practitioners will find the results helpful as a frame of reference for informing DHR-based interactive UASs and evaluating their effectiveness.

5. Intelligent User Assistance Systems to Enhance Crowd Working

Crowd work has become an important form of employment for freelancers and individuals all over the world (Kittur 2010). The concept is a digital work form based on the principles of crowdsourcing (Durward et al. 2016). Crowdsourcing is the use of Internet-based platforms to allocate individuals remotely to carry out tasks on a voluntary basis (Howe 2008). In 2017, the Leibniz Information Center for Economics estimated that 5.88% of the total active workers in the European Union considered themselves as crowd workers (Groen et al. 2017). In the USA, around 8% of adults earned a share of their income by conducting online microtasks (Gray and Suri 2019; Smith 2016). Enforced through natural growth of the digital labor market, advanced digital work forms, and the COVID-19 pandemic these numbers are expected to have increased. Especially in countries that suffer from economic pressure, crowd work gains traction since it is one of the only remaining stable and sustainable sources of income (Newlands and Lutz 2021). Also, in companies, the usage of crowd work platforms is continuously increasing. For example, in Germany, 8.3% of enterprises in the IT sector rely on services provided by the crowd for the development and production of their products (Burger 2021). The rationale for this practice is the potential to draw on specialized knowledge and to have more workforce flexibility and scalability (Burger 2021). These developments showcase the increasing value and importance that crowd work has gained for employees, companies, and the economic value chain.

A typical task in crowd work consists of the following major steps: First, crowd workers select a task on the crowd work platform which has been provided by the task provider. After the selection, the crowd workers access a presentation of the task instructions and execute it. Finally, they receive compensation for their work (Jäger et al. 2019). For microtasks like image tagging or idea generation usually take between a few minutes and one hour. These tasks are accessed via large crowd work platforms such as MTurk, Clickworker, or appJobber. Crowd workers conduct multiple, sometimes up to one hundred microtasks per day in order to generate reasonable income (Kittur 2010). This process is repetitive and mentally challenging. Together with a low income, this represents a low-quality working environment for crowd workers (Durward et al. 2020). This low-level working environment has, in turn, negative effects on task performance and the task results for the task requesters. On the one hand, it does not attract qualified crowd workers, and on the other, it reduces the motivation of the existing crowd workers.

To overcome this negative effect on crowd worker motivation and the task outcome, this work introduces intelligent UASs for the presentation of crowdsourcing task instructions. Instead of presenting information with a traditional desktop-based interface, the crowd workers are supported by an intelligent UAS that presents the task instructions and supervises the task execution (Qiu et al. 2020b). To provide a more natural interaction, the communication is done in natural language. Thus, the intelligent UAS acts as a CA (Dale 2016). This concept is called conversational crowdsourcing (Qiu et al. 2021). While microtask crowd work is structured to execute simple tasks efficiently, using conversational interfaces might be counterintuitive for efficiently conducting repetitive tasks. However, previous research on conversational crowdsourcing has shown that crowd workers perceive this interaction form as more natural and human-like, which allows for counteracting the monotonous and repetitive nature of the microtasks (Qiu et al. 2021). Further, for most of the common microtask types (e.g., creativity, user testing) the application of conversational crowdsourcing yielded equal performance results in terms of task quality and a more engaging environment for crowd workers than in traditional microtasks in previous studies (Qiu et al. 2020b).

Hence, the presentation of microtasks transforms from traditional information presentation into a human-like interaction with an intelligent UAS that acts like a task supervisor. Since crowd work does not know the traditional structure of employer and employee, this relationship between a supervisor and a subordinate allows to untap the unseen potential of leadership for crowd work (Wesche and Sonderegger 2019). In traditional hierarchical human leader-follower relationships specific leadership styles are stimulating for the followers and lead to better outcomes. Therefore, applying leadership in conversational crowdsourcing relationships with the crowd workers might be a promising alternative. Transformational leadership is the prevailing leadership theory and focuses on the stimulation of motivation and the inclusion of follower engagement (Bass et al. 2003). The application of transformational leadership capabilities by human supervisors increases the quality of the unique relationship between the leader and the follower (Fischer et al. 2017). This relationship is called leader-member exchange (LMX) (Graen and Uhl-Bien 1995; Joseph et al. 2011). An increase in LMX has a positive impact on task efficiency (Fest et al. 2021; Kovjanic et al. 2013). Further, the application of transformational leadership styles by humans on traditional crowd work task presentation has shown a positive impact on task outcome quality and crowd worker perception (Kovjanic et al. 2013). The effect of transformational leadership capabilities in conversational crowdsourcing is, however, not yet understood.

Simultaneously, the success of transformational leadership activities is influenced by the emotional capabilities of the leader (Gooty et al. 2010). Emotional capabilities may materialize in multiple forms such as a better understanding of follower emotions, more emotional sensitive language, and the display of increased understanding towards the emotions of others like empathy (Duan and Hill 1996; Pitts et al. 2012). The more perceiving and emotional sensitive to followers' emotions leaders are, the more they are able to create a comfortable and productive working environment and stimulate workers to provide better results (Cavazotte et al. 2012). In consequence, beyond the effects of transformational leadership capabilities on the followers' relationship emotional capabilities may take an additional and important role in the perception of followers also in conversational crowdsourcing. This has, however, not yet been investigated so far.

Concluding, these effects of leadership by human supervisors in traditional crowd work are positive and the usage of intelligent UASs for the presentation and supervision of crowd work tasks is stimulating. Combining the benefits of transformational leadership and conversational crowdsourcing is highly promising. However, research is scarce on how transformational leadership and emotional capabilities of intelligent UASs affect the experience of microtasks by the crowd workers and their performance in conversational crowdsourcing. Therefore, this chapter focuses on two research questions. First, what is the effect of transformational leadership capabilities of an intelligent UAS in conversational crowdsourcing on LMX, performance, and enjoyment? Second, how is the effect of transformational leadership capabilities of an intelligent UAS in conversational crowdsourcing on LMX moderated by emotional capabilities?

This chapter addresses this question by conducting a between-subject online field experiment with 239 crowd workers on the online platform MTurk. For the experiment, an idea generation task was designed in which the crowd workers had to collect innovative product ideas. This is a representative task on crowd work platforms (Cheng et al. 2020; Oppenlaender et al. 2020). They are supported in this by an intelligent UAS in the form of a CA. In the task the CA presented the task instructions and supervised the task execution. The experiment had a factorial design with four experimental treatments in which the CA either was designed with transformational leadership capabilities or not and with emotional capabilities by perceiving the crowd workers' emotional status via sentiment analysis of the textual input or not. The findings contribute to research with the transfer of the transformational leadership theory to conversational crowdsourcing. Theoretically, this work proves the relationship between

transformational leadership and LMX in a realistic conversational crowdsourcing task supported by an intelligent UAS. Further, this work provides a better understanding of its effects in the innovative digital labor environment of conversational crowdsourcing and shows the anomalies of its positive effect on the subjective experience of crowd workers' enjoyment and effectiveness and the negative effect on objective performance. Practically, the findings provide an interesting starting point for future research on the application of intelligent UASs and leadership in crowd workers' reality. Intelligent UASs provide a practical benefit by improving crowd workers' working conditions. To be implemented in reality, however, designers of intelligent UASs, first, have to overcome the deteriorating objective performance results by investigating contextual factors such as task type or AI support to stabilize or even improve the task performance for task requesters.

5.1. Theoretical Background

Conversational Crowdsourcing

Crowd work is “a digital form of gainful employment that is based on the principles of crowdsourcing in order to orchestrate an undefined mass of people via an open call on IT-facilitated platforms” (Durward et al. 2016, p. 282). The concept of crowdsourcing describes a crowd of people working on problem solving or data collection contributing to a common goal (Jäger et al. 2019). Paying a monetary compensation for the work in crowdsourcing over platforms like MTurk is, in consequence, called crowd work (Kittur et al. 2013).

Historically, crowd work gained traction in the beginning of the millennium since it provided multiple advantages for the involved stakeholders (Jäger et al. 2019). Crowd work entails three parties, the providers of the work tasks (task requesters), the online labor platforms on which the work is done, and the crowd workers who conduct the work task. Further, crowd work is classified into different categories, namely micro-work which represents short-term, routine tasks, and more complex macro-tasks that can last for multiple hours and are done by more skilled and educated workers (Kittur et al. 2013; Shafiei Gol et al. 2019).

Crowd work offers multiple benefits for the task requesters and crowd workers, such as flexibility and independency of employment (Jäger et al. 2019). However, these benefits come with unsolved challenges. Crowd work is abused for very low payments and the abduct of established work protection mechanisms such as health insurance, social security, and pensions. For example, crowd work amplifies multitasking behavior which can be origin of

inattentiveness and stress for the crowd workers (Gould et al. 2016). For the task requesters, there is a risk not to receive the adequate quality of work results. Several studies have brought attention on challenges that hinder an effective and beneficial application of crowd work (Deng et al. 2016; Oppenlaender et al. 2020; Shafiei Gol et al. 2019).

In recent years, CAs are applied to handle crowd work task execution, making it conversational (Mavridis et al. 2019). CAs are software that automatically interprets and responds to requests expressed in natural language (Dale 2016). Through their ability to interact in natural language with humans and their anthropomorphic appearance they have the potential to, both, improve the task performance but also to provide a better and more natural task experience for the crowd workers leading to higher work quality, satisfaction, and working conditions (Mavridis et al. 2019; Schuetzler et al. 2020). With the uprise of high-level capabilities through breakthroughs in natural language understanding (through models like GPT3) these developments are enforced (Kumar 2022). This concept of applying CAs to present and supervise microtasks in crowd work is defined as conversational crowdsourcing. Research showed that conversational crowdsourcing is at least as effective in task performance and can lead to a significant increase in engagement of the crowd workers (Qiu et al. 2020a, 2020b). To showcase the leadership abilities of contemporary CAs and their realization of conversational crowdsourcing, we have provided an exemplary conversation of a crowdworker with ChatGPT behaving as a leader (see Appendix C.1). In sum, conversational crowdsourcing has presented a new interaction form for crowd work task presentation and supervision. This creates an interesting alternative for innovative applications in crowd work task design.

Transformational Leadership

A dominant stream in leadership research is the transformational leadership theory (Bass and Avolio 1993). Transformational leadership has the intention to enable followers to transcend their own self-interests for a collective higher purpose or vision and to exceed performance expectations (Bass and Steidlmeier 1999; Kovjanic et al. 2013). Thereby, leaders who apply transformational leadership represent agents for an organizational change (Bass et al. 2003; Cavazotte et al. 2012; Pauleen 2003). They motivate followers by leveraging the capabilities with a stimulating vision. This vision elevates the followers' motivation and morale and transforms their self-concept (Bass et al. 2003; Bass and Avolio 1993; Shamir et al. 1993). Thereby, the followers create personal identification with the objectives of the leader – the

vision – and are motivated to achieve higher-level goals (Shamir et al. 1993). In consequence, these objectives provide the foundation for organizational change.

Transformational leadership is characterized by four leadership execution styles, namely the ability to motivate the followers (inspirational motivation), the questioning of individual approaches and stimulation (intellectual stimulation), the continuous individual discourse with followers (individualized consideration), and the provision of a feeling of putting the followers' feelings over the own feelings (idealized influence) (Bass and Avolio 1993; Howell and Hall-Merenda 1999). Multiple empirical studies and meta-studies identified a positive relationship between the application of these transformational leadership styles and the performance and efficacy of organizations (e.g., Dumdum et al. 2002). It strengthens the organizational group potency in organizations which represents the collective confidence of groups necessary to reach their goals (Guzzo et al. 1993). Through the stimulation of followers to believe in themselves and their mission, they develop competency beliefs that are a critical determinant of collective efficacy (Zaccaro et al. 1995). On the personal level, followers have a higher-level of well-being when experiencing transformational leadership (Wesche and Sonderegger 2019). In turn, drivers for the successful application of transformational leadership by leaders seem to be associated with the intelligence of the leader and certain personality traits (e.g., extraversion) (Bono and Judge 2004).

Closely connected, the emotional capabilities of transformational leaders play a large role in the effectiveness of their leadership behavior. Since they have an intensifying effect on transformational leadership (Harms and Credé 2010). Emotional capabilities manifest in many forms in transformational leadership behavior. For example, positive emotional expressions emphasize the impact of charismatic speech and the behavior of leaders (Bono and Ilies 2006; Connelly and Ruark 2010). A reason for such effect is the psychological mechanisms of contagion of the positive emotions of the charismatic leader to the followers (Bono and Ilies 2006). Previous research has shown that the strength of the elements of emotional expressions in combination with the activation of leadership has a more desirable effect (Connelly and Ruark 2010). However, shedding light on the relevance of correctly applying emotional capabilities, the direction of emotions is essential since negative emotions in an activated leadership behavior have a negative effect on followers (Connelly and Ruark 2010).

Furthermore, leaders' capability to perceive and accurately understand the followers' emotions and needs, defined as empathy, is a relevant skill for transformational leaders (Wolff et al.

2002). It precedes relevant cognitive processes and informs the behavior of the leader to individualize the focus on the followers (Wolff et al. 2002). Ergo, leaders with emotional capabilities actively identify with followers' emotions which creates an emotional experience similar to one of the followers (Sally 2000). Therefore, it is important for leaders to show a high level of emotional perception and understanding which creates an improved individual consideration of the followers (Day and Carroll 2004).

Leader-Member Exchange Theory

Beyond the sole behavior of the leader, the LMX theory has put focus on the unique dyadic relationship between leaders and their followers (Graen and Uhl-Bien 1995; Joseph et al. 2011). LMX states that “leaders establish different social exchange relationships with different followers” (Howell and Hall-Merenda 1999, p. 682). This stands in contrast to previous theories, which assumed to explain leadership as a function of personal characteristics of the leader or features of the situation (Gerstner and Day 1997). However, this implied relationship was not sufficient for explaining different follower behaviors depending on different leaders (Howell and Hall-Merenda 1999, p. 608). LMX theory states that the quality of the experienced relationship is a predictor of multiple outcomes such as work performance, mental well-being, or organizational commitment (Fischer et al. 2017). The LMX relationship shapes the work environment for the followers and influences the feelings of mutual obligation (Dulebohn et al. 2012). Looking at antecedents of LMX, Dulebohn et al. (2012) distinguish three dimensions: follower characteristics such as personality or competence, leader characteristics such as transformational leadership or leader personality, and interpersonal relationships such as perceived similarity or self-promotion. Of these, the leader characteristics revealed the strongest explanation of the variance of LMX (Dulebohn et al. 2012).

5.2. Hypotheses Development

Based on the concept of conversational crowdsourcing and the theories of transformational leadership and LMX, in this study, we investigate the effect of transformational leadership and emotional capabilities in conversational crowdsourcing. Figure 15 presents the underlying research model together with the respective hypotheses. Capabilities represent capabilities of the CA. Perceptions are perceptions of the crowd worker. We derive these hypotheses in the following.

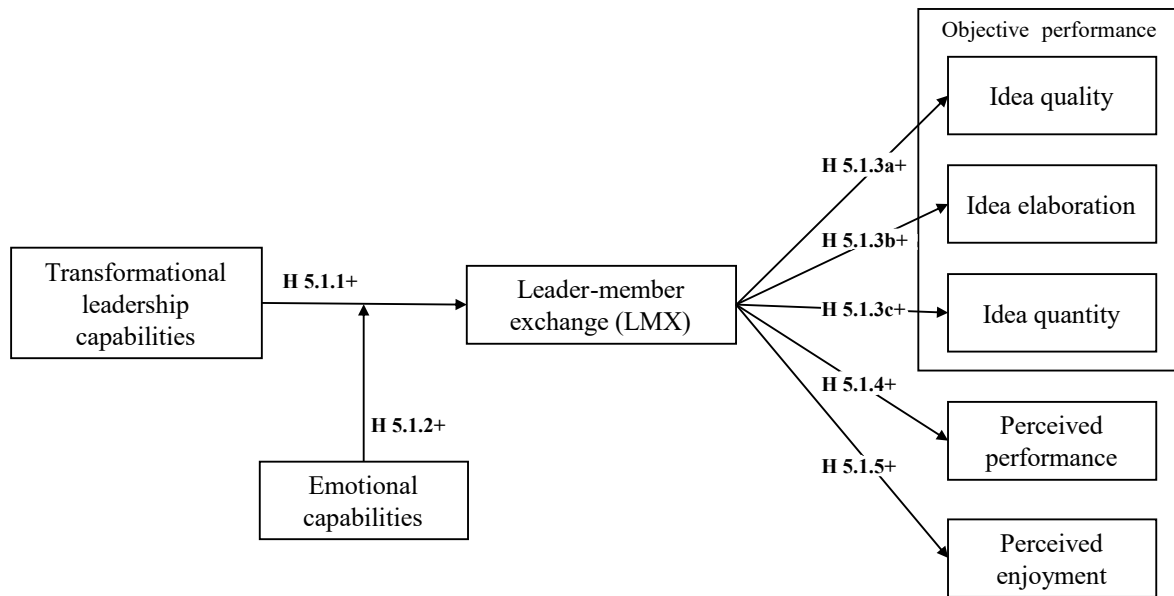


Figure 15: Research Model

Effect of Transformational Leadership on LMX in Conversational Crowdsourcing

Transformational leadership represents a formulation of a leader's vision and the identification of the followers with this vision which culminates in an adaptation of their self-concept. In human relationships, the application of transformational leadership has a strong influence on the perception of the followers. This perception is the core of LMX, a perception of the relationship between the leader and the follower (Fischer et al. 2017; Kirkpatrick et al. 1996). Multiple studies have shown the effect of transformational leadership on LMX in human leader-follower relationships (Dulebohn et al. 2012).

Crowd workers receive payment for conducting (micro-)tasks on online labor platforms on a freelance basis (Jäger et al. 2019). In conversational crowdsourcing, AI-based CAs present and supervise the task execution of the crowd workers. CAs have the capability to interact with humans via natural language and have a human-like appearance. Previous research on human interaction with CAs has shown that humans identify with them and perceive them as social counterpart (Nass and Moon 2000).

Simultaneously, with the rise of AI-based agents and software robots into work practices, the experience of leadership functions in automated agents becomes reality (Höddinghaus et al. 2021). In consequence, there have been first steps in research to investigate the effect of such functions (Wesche and Sonderegger 2019), for example, in the context of skill coaching for crowd workers (Chiang et al. 2018). Previous research showed that individuals experienced

higher integrity and transparency with automated agents in comparison to human agents (Höddinghaus et al. 2021). These are two influential factors for leadership effectiveness. Such findings on the general perception of CAs and on human perceptions of automated leadership suggest that the effect of transformational leadership also holds for the interaction of humans with leadership in CAs. In crowd work, a CA that formulates a vision and provides motivational suggestions as transformational leadership capabilities might have a positive effect on the crowd worker. Based on these findings, we assume that transformational leadership capabilities in CA in crowd work increase the perceived quality of LMX. Therefore, we transfer these findings to the context of conversational crowdsourcing and derive the following hypothesis:

Hypothesis 5.1.1: A conversational agent with transformational leadership capabilities in conversational crowdsourcing increases leader-member exchange compared to a conversational agent without transformational leadership capabilities.

Influence of Emotional Capabilities on Transformational Leadership

Transformational leadership is influenced by multiple characteristics of the leaders such as personality, context, and the leader's capabilities. One important asset for leadership effectiveness are the emotional capabilities, the abilities of leaders to understand, handle, and apply emotions (Ferris et al. 2009). A potential reason for the importance of emotional capabilities in leadership is that emotions are one of the core indicators of a positive or negative relationship between the leader and the follower (Graen and Uhl-Bien 1995). The capability to perceive and understand the emotions of the followers and to dynamically react and regulate them is highly beneficial in order to create a stable and valuable relationship between the leader and the follower (Day and Carroll 2004; Gooty et al. 2010). This effect increases the more relational aspects are present in the LMX relationship and less rational the relationship is (Ashkanasy and Daus 2002). Further, emotional capabilities by a CA materialize in multiple ways, for example, in the form of increased empathy. Empathy is the "process of humanizing objects, of reading or feeling ourselves into them" (Duan and Hill 1996, p. 1). This emotional reading and showing emotional understanding of the situation of the crowd worker is similar to the perception of social support. Social support for crowd workers, in turn, leads to an increase in crowd workers' engagement (Ihl et al. 2020). Reasons for this development are the creation of an identification of crowd workers with their task and the idea behind it and the enhanced meaningfulness of their work (Ihl et al. 2020). Harms and Credé (2010) demonstrated the

importance of this effortful connection between emotional capabilities and transformational leadership capabilities in their meta-analytical review.

Combining the findings on the relationship between leaders' emotional capabilities with transformational leadership and the transfer of the effect of transformational leadership to automated CAs, we derive that CAs that are capable of showing emotional capabilities like empathetical behavior emphasize the positive impact of transformational leadership capabilities on the relationship between the CA as a leader and the crowd worker in conversational crowdsourcing. We hypothesize:

Hypothesis 5.1.2: A conversational agent leveraging emotional capabilities in conversational crowdsourcing positively moderates the relationship of a conversational agent with transformational leadership capabilities and leader-member exchange compared to a conversational agent without emotional capabilities.

Effect of LMX on Objective Performance and Crowd Worker Perceptions

According to LMX theory, a high-quality leader-follower relationship perceived by the follower increases the intrinsic motivation to invest more effort in the work at hand (Wayne et al. 2002). High quality of LMX represents a high valuation of the working conditions by the followers. Also, it represents a higher level of trust in their supervisors. The followers feel the need to reciprocate this beneficial treatment by increasing their work effort. An increased investment of effort correlates with an improvement in outcome quality. This holds, in particular, in crowd work microtasks which do not require special education or prior knowledge (Kittur et al. 2013). Herein, microtasks in crowd work do not differ from physical work.

In conversational crowdsourcing, a CA may act as a leader who is fostered through its human-like appearance (Qiu et al. 2021). Previous knowledge has shown that such a human relationship also establishes with a CA, in particular, if the CA is accepting, understanding and non-judgmental, which are typical transformational leadership characteristics (Skjuve et al. 2021). Through these characteristics, we argue that the positive effects of LMX on outcome performance are also valid in the context of conversational crowdsourcing. This transfer to conversational crowdsourcing means that the task performance of crowd workers in microtasks improves. Hence, we believe that higher quality of LMX leads to an increase in objective performance of conversational crowdsourcing tasks. We formulate the following hypothesis:

Hypothesis 5.1.3: The quality of the leader-member exchange perception is positively related to objective performance of crowd workers in conversational crowdsourcing.

In crowd work, two dimensions play a role in the performance outcomes of the task crowd workers are performing. First, the crowd workers perform their work and this can be measured. For example, for labeling tasks this is the number of correct labels set. This represents the objective performance and is, in particular, interesting for the crowd work task requesters.

Second, the crowd workers' perspective is relevant. Crowd workers assess the outcome of their work. Their assessment has a strong influence on their self-perception and identification with the task (Durward et al. 2020). A better relationship between the leader and the follower (i.e., LMX) positively stimulates this assessment due to higher trust and a facilitated process of identification (Kittur et al. 2013). Following this idea, crowd workers that invest more effort into their work due to a better LMX assume a higher effectiveness of themselves. This higher perceived effectiveness is equal to their subjective performance perception. Therefore, high quality LMX promotes the perceived personal value crowd worker are drawing from their work. We hypothesize:

Hypothesis 5.1.4: The quality of the leader-member exchange perception is positively related to the subjective performance in form of perceived effectiveness by the crowd worker in conversational crowdsourcing.

Hypothesis 5.1.3 and hypothesis 5.1.4 relate to objective and subjective performance of the task. Simultaneously, a relevant aspect for the crowd worker perspective is the enjoyment crowd worker experience while conducting the task. Together with the perceived effectiveness it is a factor for influencing perceived self-efficacy of crowd workers (Sitzmann and Yeo 2013). The experienced task enjoyment determines the intrinsic task motivation (Sansone and Morgan 1992). In creative task processes, task enjoyment is described as the degree of having fun and satisfaction during a creative process and the absence of annoyance and frustration (Dahl and Moreau 2007). A high level of task enjoyment makes humans more likely to engage in an activity in the future (Sansone and Morgan 1992). In crowd work, this means that it is easier for task requestors to attract crowd workers in future tasks. An improvement in the relationship of the leader and the follower, a higher LMX, leads to an increase in well-being at the workplace (Sparr and Sonnentag 2008), in particular in the form of control and satisfaction with the task at hand (Dulebohn et al. 2012). Since we assume that these findings are also valid for the relationship between an AI-based CA and crowd workers, we transfer them to the context of

conversational crowdsourcing. Hence, we argue that increased LMX between a CA as leader and a crowd worker as follower increases the enjoyment of the crowd worker during the task execution. We hypothesize:

Hypothesis 5.1.5: The quality of the leader-member exchange perception is positively related to the perceived enjoyment of a crowd worker in conversational crowdsourcing.

5.3. Methodology

In this study, we conducted an online experiment as a conversational crowdsourcing task to generate innovative ideas guided by a CA on the platform MTurk and a second crowd work task to assess the quality of the generated ideas from the first task.

Conversational Crowdsourcing Experiment

Participants

We recruited 256 participants via MTurk. The payment for experiment participation was divided into a fixed and variable share. Each participant received a fixed amount of \$4.50 for task completion and a variable amount of up to \$0.30 depending on the total number of ideas generated. The payment corresponds to the standard average wage level in the country of the study's institutions (\$13.0 per hour). To ensure the collection of high-quality data, we rigorously excluded 17 participants who failed one of multiple attention checks (one reverse-coded item, one control item instructing participants to check the second leftmost answer option, an introductory question regarding the nature of the task). The final sample consisted of 239 participants with an average age of 38.85, of which 111 participants were females (46.44%) and 128 males (53.56%).

Experimental Design and Procedure

To test our hypotheses, we conducted an online field experiment with four experimental treatment conditions. To have a natural crowdsourcing task, we created a human-intelligence task (HIT) for the creation of design ideas for a product. The HIT was guided by a CA with the name "Crowy". We designed the CA with four conditions of either having transformational leadership capabilities or not and either having emotional capabilities or not (2x2 factorial design). Participants were randomly assigned to one of the treatment conditions (between-subjects design). The experiment is a natural field experiment, according to Harrison and List

(2004), due to the participant pool, information, task, stakes, and environments which are typical for crowd work on MTurk.

The experimental procedure is presented in Figure 16. After starting the HIT, the participants were informed about the task objective, the generation of innovative ideas for the product, the payment scheme, and the task duration. We framed the task as a task for finding new product ideas for travel cases on behalf of a research and development department of a travel cases manufacturer. We chose travel cases since they are familiar to the majority of the crowd worker population and applicable for product research by enterprises. This was followed by demographic questions. Following the general introduction, participants were randomly assigned to one of the treatment conditions and exposed to the task interface. It consists of the chat interface on the left side and information about task, time, and generated product ideas on the right (see Appendix C.2). In the task interface, participants started to communicate with the CA that introduced itself, provided a detailed description of the task, and checked for correct task understanding (i.e., asking “*For what type of product should you find new product application ideas?*”). Afterwards, they started the task execution. The CA asked two questions to assess the emotional state of the participant (i.e., “*How do you feel today?*”, “*Do you like travel cases?*”). Then, the CA asked the participants to start the task and, when the crowd workers were ready, the CA started the idea generation phase in which the participants were able to enter ideas. This phase lasted for ten minutes. During the idea generation phase, participants were able to share their ideas with the CA via the text interface. Ideas were presented in list form next to the chat. After ten minutes, the CA finished the task and provided the participant with a code to continue. Subsequently, the participants answered a questionnaire and received a debriefing about the experimental nature of the task.

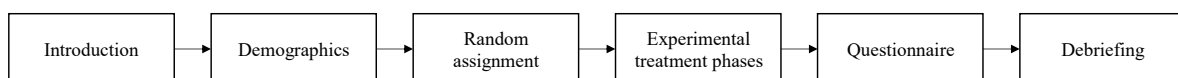


Figure 16: Summary of the Experimental Procedure

Experimental Task

Our goal was to conduct a field experiment to investigate the experience of real-world crowd workers in their natural environment (i.e., on MTurk) (Harrison and List 2004). Besides online freelancing for larger projects (e.g., programming, proof-reading), a large share of crowd work is done via microtasks (Durward et al. 2020). A common microtask form is creative idea generation or ideation tasks (Oppenlaender et al. 2020). Those tasks are characterized as “open-

ended tasks and activities in which users develop new ideas” (Kerne et al. 2014, p. 1). Using crowd work to benefit from the creativity of the crowd workers is promising and widely applied due to the easy access to a larger number of diverse individuals (Oppenlaender et al. 2020). Therefore, to evaluate the effect of transformational leadership capabilities in conversational crowdsourcing, we decided to develop a creative ideation task based on the task design of Kovjanic et al. (2013), who investigated the relationship between a human leader and followers in their crowdsourcing experiment. The experimental task was to generate as many ideas for travel cases as possible within ten minutes. To ensure a minimal idea quality, idea descriptions had to contain at least five words. We, further, required the participants to provide a minimum of three ideas to successfully complete the task to prevent participants from simply letting time pass.

Experimental Treatment Phases

The experience of the experimental treatments took place in *three experimental treatment phases (ETP)* (see Figure 17). In the first *ETP (ETP1)*, the participants read through two informative screens that contained an introduction text for the travel case innovation task by the CA but not in a conversational form (see Appendix C.2 for details). Afterwards, they entered *ETP2* in the form of a conversational chat window. In this second phase, an additional conversational introduction took place. In the third *ETP3*, the participants kept on conversationally interacting with the CA to execute the ideation task. In this phase, both the submission of the ideas and interaction with the CA took place. We separated the treatment experience into these phases since the interaction with a CA is different from the interaction with traditional desktop-based information. The amount of information that was previously presented via pages with information overloads the textual conversational interaction in a chat window. Therefore, we decided to send an informative page as an address by the CA to the crowd worker and continued with the conversational interaction in *ETP2*. With this approach of an initial textual address by a leader as part of the experimental treatment, we followed Kovjanic et al. (2013) and Felfe and Schyns (2006). We present an overview of the transformational leadership and emotional capabilities of the three *ETPs* in Appendix C.3, as well as exemplary treatment capabilities of the *ETP* in Appendix C.4.

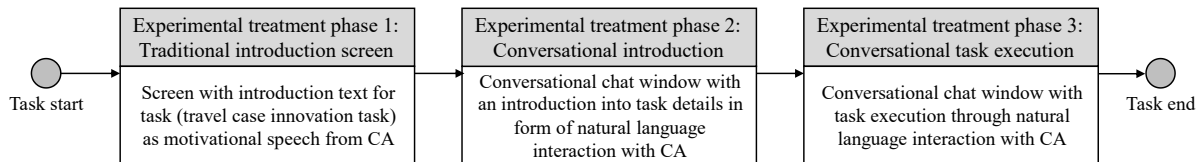


Figure 17: Summary of the Three Experimental Treatment Phases

Experimental Treatment Conditions

We employed a between-subjects full factorial design with two factors: transformational leadership capabilities (presence/absence) and emotional capabilities (presence/absence). The four resulting treatments are:

- (1) Baseline treatment: The CA has no leadership and no emotional capabilities.
- (2) Leadership treatment: The CA has transformational leadership capabilities only.
- (3) Emotional treatment: The CA has emotional capabilities only.
- (4) Full treatment: The CA has transformational leadership and emotional capabilities.

We implemented a CA with the name “Crowy” with a common CA development framework (Microsoft Bot framework). The CA led the crowd workers through the ideation task. After the welcome screen, it introduced itself to the participants by its name. From that moment on, all information was provided to the participants by the CA. For the CA design with transformational leadership capabilities, we had to choose the gender of the CA. However, to the best of our knowledge, there are no established scientific findings in leadership research with regard to the role of gender of CAs and its relation to the strength of leadership. To solve this challenge, we determined the CA gender dynamically to match the gender of the participant. Previous research has shown that interacting with an avatar of the same gender yields higher persuasiveness (Guadagno et al. 2007). Based on the answers to the demographic questions before the treatment experiment, the CA adapted to the crowd workers’ gender. For example, the CA was female for female crowd workers. The gender was also represented in the identifying image of an avatar for the CA (see Figure 18). To maintain comparability, the avatar wore identical clothing (i.e., jumper, shirt) and had similar facial features and characteristics (i.e., hair color).



Figure 18: Representation of the Female and Male Conversational Agent

Transformation leadership includes four leadership execution styles. This section describes these four transformation leadership execution styles in detail. Three of those serve as the foundation for the experimental treatment conditions. According to Bass et al. (2003) the execution styles consist of the following in detail:

Inspirational Motivation: Leaders create team spirit by providing enthusiasm and optimism combined with a meaning and a continuous challenge of the followers' work. They achieve that by envisioning an attractive future state to reach, which the followers overtake for themselves in the course of the identification with the leader's goals.

Intellectual Stimulation: Transformational leaders pursue to stimulate the thinking process of their followers to reach innovative and creative solutions. Therefore, they continuously question assumptions of the status quo and reframe existing problems to approach them in a new way. Additionally, in transformational leadership follower mistakes are not the origin of public criticism.

Individualized Consideration: A core of transformational leadership is the focus on the follower as an individual. Therefore, leaders consider and mention each individual's need for achievement and growth. Leaders rather act as mentors than as pure assessors and have the goal to develop the potential of their followers. They support this via learning opportunities and by considering each individual's strengths and weaknesses.

Idealized Influence: The idea of a transformational leader is that such leaders are admired by their followers. They achieve this by the presentation of different values in their actions, such as putting the followers' needs over their own, sharing common risks together, and acting according to common ethical principles. Followers often perceive this as charisma which creates trust and respect. It lets the followers want to identify with the leader.

To introduce transformational leadership capabilities into the experimental treatment, we operationalized the three styles of the transformational leadership theory: individualized consideration, intellectual stimulation, and inspirational motivation (Bass and Avolio 1993;

Bass and Steidlmeier 1999). We decided against the operationalization of idealized influence. Idealized influence in the leader is defined as “charisma which is envisioning, confident, and sets high standards for emulation” (Bass and Steidlmeier 1999, p. 187). In our creativity microtask, the CA has limited short-term interactions with the crowd workers. Since the provision of charisma is highly complex, it is difficult to operationalize it in a CA.

The three leadership styles were selectively applied in different forms throughout the experimental task by the CA. For individualized consideration, the CA showed individual attention to the crowd worker (i.e., “*I have checked your ideas...*”), for inspirational motivation the CA tried to individually motivate the crowd worker (i.e., “*We will be successful!*”), and for intellectual stimulation the CA aimed to provide support for solution seeking (i.e., “*Think about what you like.*”). A detailed description of the application form and timing of the leadership styles is presented in Appendix C.3.

We operationalized the emotional capabilities with a system-driven adaptation of the messages. Therefore, we measured the sentiment in the answers to the two small-talk questions in the beginning of the chat interaction between the crowd workers and the CA. For this assessment, we used a sentiment analysis that abstracts discrete emotions on a continuous scale from negative to positive (from -1 to 1). For this, we used the sentiment analysis tool VADER (Hutto and Gilbert 2014). Depending on the positive or negative sentiment of the answers, the CA changed its answers (e.g., “*That's nice to hear*” vs. “*Well, then let's forget the rest of the day!*”) and used emojis (e.g., “*:-*”).

Measurement Variables

LMX: To measure the quality of the exchange relationship between the CA and the participant, we adapted the LMX-7 questionnaire. The LMX-7 is a well-established questionnaire that measures the relationship between the leader and the follower (van Breukelen et al. 2006). Due to the special case of leadership in conversational crowdsourcing and the short-term interaction in a microtask, we removed three items that address a traditional long-term work relationship (e.g., “*My leader would bail me out at his/her expense*”). We adapted the remaining four items to the experimental scenario. All items used for the experiment and the following scales are provided in Appendix C.5.

Objective Performance: To complement the subjective perception of participants via survey, we conceptualized the objective performance of the participants in an idea generation task with

three measures: quality, elaboration, and quantity of the ideas submitted. Since idea quality is complex to assess objectively, we followed a common approach in creativity research for idea quality assessment (e.g., Wu and Bailey 2021). We conducted a second study with a new cohort of crowd workers (376 participants) in which multiple crowd workers assessed the quality of each idea. Their assessments were then averaged. We present the details of this evaluation in the following section (Section *Idea Quality Evaluation Study*). While this approach is not entirely objective by nature since humans make a judgment, we follow established methods and assume a high level of objectivity. In the following, we will, therefore, speak of objective performance. Second, we collected the elaboration of each idea in terms of written letters per idea. The more detailed an idea is described, the easier the task requester can process it for further usage. A higher elaboration of the idea is an indicator of better performance. Third, we assessed idea quantity and counted the number of ideas produced by each crowd worker. To assure uniqueness, two authors checked the ideas independently and removed redundant ideas.

Perceived Performance: We measured the subjective performance perceptions of the participants with a four-item scale of perceived effectiveness by Dennis (1996) and Dennis et al. (1996).

Perceived Enjoyment: In addition to performance, we measured the perceived enjoyment of the crowd workers during task execution to analyze the enjoyment of the participants when conducting conversational crowdsourcing. Therefore, we used a three-item scale by Füller et al. (2009).

Manipulation Checks: To ensure that the manipulation of the transformational leadership capabilities was effective, we collected the perceived leadership effectiveness of the CA with an adapted scale by van Knippenberg and van Knippenberg (2005). For emotional capabilities of the CA, we collected the perceived empathy of the CA with an adapted scale by Watson et al. (1998) since empathy is a valid signal for increased emotional capabilities. We removed two items to maintain suitability for the experimental scenario.

Controls: We controlled for demographic differences in age, gender, and education, as well as experience with crowd work and CAs.

Idea Quality Evaluation Study

To assess idea quality, we conducted an additional study in which each idea was evaluated by exactly three individual crowd workers. Through this approach, we reduced the influence of personal attitudes and perceptions.

Participants

For this study, we recruited 376 participants via MTurk. Quality assessment is a regular task on MTurk and crowd workers are familiar with this task type. Participants received a fixed payment of \$2.00. We ensured that none of the participants had previously participated in the first task by excluding the respective crowd workers' IDs. To maintain idea quality, we excluded 42 participants who failed one of the two attention checks (i.e., two instructional control questions that requested not to answer them). The final data set consisted of 334 participants with an average age of 39.5, of whom 139 (41.62%) were female and 195 (58.38%) were male.

Procedures

First, the participants received an introduction that explained that the participants had to evaluate the quality of innovative travel case ideas. The introduction specified that the ideas were not limited to any area or domain. Second, participants received a tutorial for the assessment of the idea quality and the relevant quality dimensions. For the quality dimensions, we followed Dean et al. (2006) that defined four idea quality dimensions: novelty, feasibility, relevance, and specificity. Subsequently, we aggregated them into one idea quality measure. The items for feasibility and relevance are derived from the definition by Dean et al. (2006). For novelty and specificity, we used items by Blohm et al. (2011). This resulted in a four-item scale for idea quality (i.e., Please rate the extent: (1) "to which the idea is novel", (2) "to which the idea can be implemented easily (considering costs, time, infrastructure, legal issues, etc.)", (3) "to which the idea's utility is clearly described", (4) "to which the idea clearly applies to the stated problem"). Third, the participants received an example of an idea description to familiarize themselves with the idea complexity. After this introduction, the idea quality evaluation began. For the evaluation, the participants received 16 ideas randomly selected from the idea pool and evaluated them on the four-dimensional scale. After evaluating all 16 ideas, participants were asked to provide demographic information and the study ended.

5.4. Results

An initial descriptive analysis reveals that the treatments with transformational leadership capabilities have higher means for LMX, idea quantity, perceived effectiveness, and perceived enjoyment compared to treatments without them (see Table 12 for descriptive values). However, we also see lower values for idea quality and elaboration with the transformational leadership treatments. The treatment with emotional capabilities reveals multifaceted results since the mean for LMX, idea quantity, idea elaboration, and perceived effectivity is higher than in the baseline treatment but lower for idea quality and perceived enjoyment. Further, the mean of the treatment with leadership and emotional capabilities is between the mean of the treatments with single capabilities.

Variables		Experimental treatment conditions			
		Baseline characteristics	Transformational leadership capabilities	Emotional capabilities	Leadership and emotional capabilities
LMX	Mean	3.119	5.180	4.000	5.072
	SD	1.493	1.394	1.719	1.429
Objective performance: Idea quality	Mean	4.807	4.608	4.736	4.629
	SD	0.682	0.813	0.741	0.788
Objective performance: Idea elaboration	Mean	114.594	100.524	117.971	111.434
	SD	69.201	53.268	72.973	62.53
Objective performance: Idea quantity	Mean	6.424	7.264	6.734	6.896
	SD	3.292	2.811	3.02	2.477
Perceived effectiveness	Mean	5.441	5.599	5.465	5.635
	SD	0.898	1.064	1.024	1.191
Perceived enjoyment	Mean	5.147	5.471	5.135	5.333
	SD	1.429	1.515	1.46	1.602

Table 12: Mean and Standard Deviation in the Experimental Treatment Conditions

To test our hypotheses, we chose structural equation modeling (SEM). Since the underlying data did not meet the assumptions of normality, we applied partial least square SEM (PLS-SEM) rather than covariance-based SEM. PLS-SEM allows to model latent constructs under the condition of non-normality with small to medium sample sizes (Jannoo et al. 2014; Wang et al. 2011). Standard criteria suggest adequate reliability and validity of the measurement model (see Appendix C.6 for details). A test for common method variance among latent constructs shows that this is not an issue in our data (see Appendix C.7 for details). To complement our findings, we also computed the initially planned covariance-based SEM

approach with R studio (v.2021.09.2) and the lavaan package (v.0.6-10), which confirmed our results.

The experiment design includes the manipulation of the CA’s transformational leadership and emotional capabilities. We tested the success of the leadership manipulation by asking how much the participants perceived the CA as a leader (with the construct perceived leadership effectiveness, e.g., “Crowy is a good leader.”). The results show that the CA was significantly more perceived as a leader in the treatment with leadership characteristics. Further, we tested the emotional capabilities manipulation with the perceived empathy (e.g., “Crowy gave you individual attention.”). The analysis shows that the manipulation was effective (see Appendix C.8 for details).

Evaluation of Structural Model

After verifying the measurement model validity, we proceed with analyzing the structural model, including the CA’s capabilities. Figure 19 shows the estimation results of the structural model and Table 13 provides a summary of the hypotheses evaluation. We used 5,000 bootstrapping resamples to estimate the significance of the path coefficients.

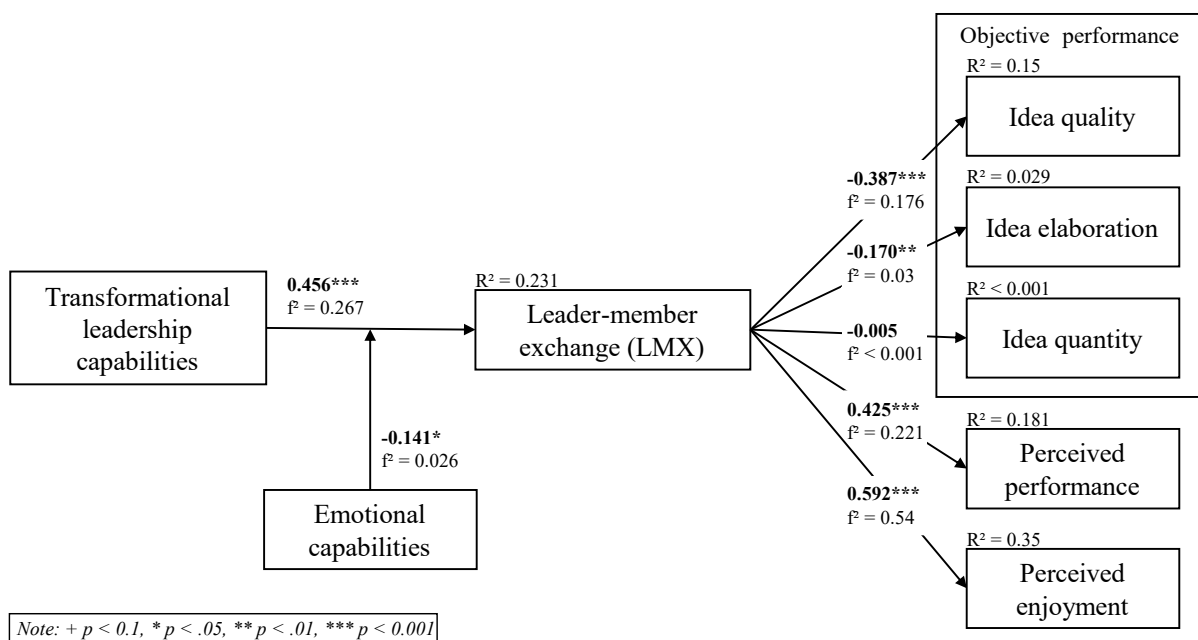


Figure 19: Structural Equation Model of the Research Model

As indicated by the descriptive analysis, transformational leadership capabilities by a CA have a positive effect on LMX ($\beta = 0.456$, $f^2 = 0.267$, $p < 0.001$). Following Cohen et al. (1998) we report a medium effect size. Thus, we find support for hypothesis 5.1.1. The positive effect of transformational leadership capabilities in a CA on LMX is moderated by the effect of

emotional capabilities in a CA. Contrary to our expectation, emotional capabilities in the CA have a small negative moderating effect on the relationship between transformational leadership capabilities in the CA and LMX ($\beta = -0.141, f^2 = 0.026, p < 0.05$). Thus, hypothesis 5.1.2 is not supported. Not being part of our research model, we also analyzed the relationship of emotional capabilities by a CA on LMX. To complement our analysis, we see a positive direct effect ($\beta = 0.116, f^2 = 0.017, p < 0.05$).

Subsequently, the influence of LMX on the different dimensions of objective performance is predominantly negative. The effect of LMX on idea quality shows a medium significant negative effect on idea quality ($\beta = -0.387, f^2 = 0.176, p < 0.001$) and a small significant negative effect on idea elaboration ($\beta = -0.170, f^2 = 0.030, p = 0.008$). For the effect of LMX on idea quantity we find no significant effect ($\beta = -0.005, f^2 < 0.001, p = 0.937$). Therefore, we have to reject the hypothesis 5.1.3 since some of the subdimensions (i.e., quality, elaboration) do not show the hypothesized effect. In contrast, for the effect of LMX on the subjective performance perception, we report a medium significant positive effect for perceived effectiveness ($\beta = 0.425, f^2 = 0.221, p < 0.001$). We report support for hypothesis 5.1.4. Finally, the results show a strong positive effect between LMX and perceived enjoyment of the crowd workers during the task ($\beta = 0.592, f^2 = 0.540, p < 0.001$). Therefore, we find support for hypothesis 5.1.5.

Hypotheses	Results	Findings
H 5.1.1	Supported	Providing CAs with transformational leadership capabilities in crowd work increases LMX.
H 5.1.2	Not supported	Providing CAs with emotional capabilities in crowd work negatively moderates the relationship of CAs with transformational leadership capabilities on LMX.
H 5.1.3	Not supported	(a) Higher quality of LMX decreases objective performance in form of idea quality in conversational crowdsourcing. (b) Higher quality of LMX decreases objective performance in form of idea elaboration in conversational crowdsourcing. (c) Higher quality of LMX has no significant effect on objective performance in form of idea quantity in conversational crowdsourcing.
H 5.1.4	Supported	Increased LMX increases the subjective performance in form of perceived effectiveness by the crowd worker in conversational crowdsourcing.
H 5.1.5	Supported	An increase in LMX will increase the perceived enjoyment of the crowd workers in conversational crowdsourcing.

Table 13: Summary of Hypotheses Testing

5.5. Discussion

The primary goal of our study was to investigate the effect of transformational leadership and emotional capabilities in conversational crowdsourcing. Based on the transformational leadership and LMX theory, our results show that a CA's transformational leadership capabilities lead to an improved relationship between the CA and the crowd workers. This, in turn, resulted in higher subjective perceptions of effectiveness and enjoyment of the crowd worker. Interestingly, objective performance remains unaffected (i.e., idea quantity) or deteriorates (i.e., idea quality, elaboration) depending on the objective performance measure applied. Emotional capabilities did not emphasize but mitigate the effect of transformational leadership capabilities. From our findings, we derive several contributions to research and practice.

Contribution to Theory

Positive Effect of CA's Transformational Leadership Capabilities on LMX

To the best of our knowledge, our study is the first to investigate and show the positive effect of transformational leadership capabilities in conversational crowdsourcing agents on the relationship between the crowd worker and such agents (i.e., LMX) as perceived by the crowd workers. Thus, we contribute knowledge on this interesting relationship in conversational crowdsourcing research. Noteworthy, we are able to support the existing theoretical findings from human relationships between leaders and followers (Dulebohn et al. 2012; Fischer et al. 2017). Hence, they are transferrable to leadership relationships between humans and CAs. Thereby, we contribute to the theory of transformational leadership and LMX by showing that the effects are also valid in the context of CAs and conversational crowdsourcing in particular.

Negative Moderation Effect of Emotional Capabilities

Our results show a negative moderation effect of emotional capabilities on the relationship between transformational leadership capabilities and LMX. That means with an increase in emotional capabilities in the CA, we observe a weaker effect of transformational leadership capabilities on LMX. This, however, stands in contrast to the theoretical foundation of transformational leadership theory and the expected outcome. While transformational leadership capabilities in the form of motivation and a vision positively impacted the perception of the relationship between the CA and the crowd worker, additional emotional capabilities led

to a weaker positive impact on this relationship. Thus, the crowd workers did not appreciate if the CA had leadership and emotional capabilities. A potential interpretation of these results might be that the human crowd workers valued the positive impact of being guided with a vision but disliked an emotional CA. It appears that crowd workers perceive the leadership guidance support by a CA as a functional characteristic and, henceforth, as valuable for their task at hand. While we hypothesized that emotional capabilities provide a form of social support that increases worker engagement (Ihl et al. 2020) this support had a negative influence. Social support on the emotional level is a human capability. At the same time, crowd workers are aware that the CA is not human. This implies that the crowd workers did not trust a CA to play the role of a human emotional leader and to provide social support. The CA became less authentic and there was no additional functional value of the emotional capabilities for the crowd workers. This might have diminished the transformational leadership effect on the relationship between CAs and crowd workers.

Contrasting Effect of LMX on Objective and Subjective Performance

The performance variables of our study yielded unexpected results. According to LMX theory, an increase in LMX by the follower leads to an increase in task performance, task enjoyment, and emotional commitment (Dulebohn et al. 2012; Gerstner and Day 1997). While the crowd workers reported an increase in enjoyment, our findings, however, show a significant negative effect on two objective performance variables, namely idea quality and idea elaboration. In detail, the crowd workers that experienced transformational leadership capabilities had a better perception of the relationship with the supervising CA but were less performant in creating high-quality ideas and in providing more detailed descriptions. This finding in the context of conversational crowdsourcing is interesting since it contradicts the theoretical foundation of LMX theory in human relationships.

Further, we want to shed light on the task of our study as a specific use case of crowd work. Creativity tasks such as the idea generation task we employed are common on crowd work platforms (Oppenlaender et al. 2020). However, they are also a special form of crowd work tasks which distinguish from matching or labeling tasks in their cognitive and emotional requirements for the crowd workers (Oppenlaender et al. 2020). Previous research showed that a higher quality of LMX perceived by the followers increases the autonomy of the followers together with their creativity (Volmer et al. 2012). That means in a traditional working relationship between human leaders and followers a better and individual relationship leads to

more and better ideas in a creativity task (Wong et al. 2021). In our experiment, however, with a CA with transformational leadership capabilities and a human crowd worker, these findings are reversed, and we receive significant negative results in terms of idea quality and elaboration with higher quality of LMX. An intuitive approach would be to ask whether a microtask interaction with a CA allows for the establishment of a leader-member relationship. The results of our manipulation check by asking for specific items (such as “*Crowy was a good leader*”) refute this thought. Overall, this shows that results from traditional transformational leadership theory and creativity are not directly transferrable to conversational crowdsourcing.

In consequence, there must exist a negative effect that counteracts the positive effects of increased LMX and that lowers the idea quality and quantity effectively. A potential explanation for this negative development of the objective performance might be due to the fact of a creativity task in the context of conversational crowdsourcing. A CA with leadership capabilities in crowd work provides a vision and motivation and guides through the crowd worker task. Inevitable by definition, guidance and leadership in crowd work tasks create a directive as well as mental constraints and limits for the crowd workers. Following this line, previous work has shown that helping behavior provided by a supervisor in crowdsourcing communities had negative effects on the performance (Dissanayake and Sridhar 2021). Further, in creativity, it is, however, important to have mental freedom and the absence of constraints to unlock ideation flow (Halbesleben et al. 2003). In crowd work, findings showed that crowd workers may perceive any form of virtual feedback as surveillance and this perception, in consequence, hurts their creative performance (Wong et al. 2021). Simultaneously, findings from team research found evidence that standardization and a certain level of constraints during the creative activity come with benefits for individuals since it reduces ambiguity and complexity (Dahl and Moreau 2007; Gilson et al. 2005). Besides the potential negative effect related to the creativity boundaries another explanation might lie within the characteristic of leadership. In particular, transformational leadership behavior is often expressed in additional interaction between the supervisor and the supervisee (de Vries et al. 2010). This might have increased the cognitive load of participants and simultaneously led to negative effects on the crowdworkers’ performance. In sum, we cannot draw a clear conclusion about the reasons behind our results. However, when comparing the positive effect of LMX in human relationships and of constraints on creativity with the negative effects, we observe that the negative effects of transformational leadership capabilities outweigh the positive effects on creativity.

Practical Implications

Crowd worker Perspective: Positive Impact on Crowd Work Conditions

At first, we want to highlight the positive effect of the transformational leadership capabilities of CAs in crowd work on the general experience of the crowd workers. From a practitioner's perspective, this is highly valuable since it improves the overall situation of the most vulnerable part of the crowd work value chain, the individual crowd workers. In traditional working environments such as manufacturing or health services, over the years of centuries, several worker support initiatives and institutions have established themselves. Workers are supported and protected by law for a limitation of working hours, holidays, and working conditions. This increases the status of the individual worker and their valuation and contributes to their general perceived value and self-confidence. Our study results showed that the workers perceived higher effectiveness of their work and more enjoyment during the execution. Both are important psychological drivers for perceived self-efficacy (Sitzmann and Yeo 2013). Self-efficacy, in turn, creates optimism and self-confidence (Karademas 2006). In sum, from a social perspective of the crowd work workforce, the application of conversational crowdsourcing improves the working conditions and is therefore desirable.

Task Requester Perspective: Ambivalent Findings for Conversational Crowdsourcing Implementation

On the flip side, our ambivalent results do not allow to derive the unequivocal application of transformational leadership capabilities in conversational crowdsourcing for task requesters such as companies. Our results reveal a significant negative effect on objective performance when applying transformational leadership and emotional capabilities. Simultaneously, it is highly beneficial for task requesters to have content and satisfied crowd workers who are willing to work for the requesters over a longer period. While crowdsourcing tasks are usually short, crowd workers decide what tasks of which requester they accept. Therefore, it is important if crowd workers are willing to conduct tasks for the requesters again. However, we believe that there are certain reasons for these results and the practical implication is limited to the current level of knowledge. Several factors play a role in the performance of crowd workers in general and the effect of transformational leadership capabilities on the creative task performance in conversational crowdsourcing in particular. For example, the task introduced by conversational crowdsourcing provides a certain frame for leadership by a CA. The different elements of transformational leadership might have adverse effects. When transferring our

results to practice, we should also keep in mind the lack of long-term empirical evidence for the effect of CAs' transformational leadership capabilities in case of individual crowd workers repeatedly interact with the CA. For one, the positive effects might wear off over time. On the contrary, the relationship might grow stronger over time, increasing the positive effects. Further, habitual interaction might lead to higher objective performance. However, these are aspects to keep in mind, but we do not have empirical evidence of such effects yet.

To articulate practical implications, based on our results, we find no evidence that emotional capabilities have a positive effect, and thus, our results suggest not to implement emotional capabilities. For transformational leadership capabilities, we cannot make a recommendation for or against the use of conversational crowdsourcing for companies in their value chain at this time. We still have to dig deeper and further investigate the root causes for the objective performance effects. Considering the positive and negative practical implications jointly, we derive as the bottom line that there are positive factors of transformational leadership in conversational crowdsourcing which are valuable and should be retained. Simultaneously, to retain the positive aspects, it is necessary to lift the objective performance of conversational crowdsourcing.

Limitations and Future Research

Although we followed a rigorous experimental procedure, limitations apply to our work. First, we used a creativity task to assess the effect of transformational leadership and emotional capabilities in conversational crowdsourcing. While creativity tasks are prominent tasks on crowd work platforms (Oppenlaender et al. 2020) there are multiple task types on crowd work platforms, e.g., labeling, programming tasks (Kittur et al. 2013). Future work should investigate the identified effects in different conversational crowdsourcing tasks.

Second, microtasks are per definition short-term working environments and the interaction of crowd workers with the CA in our experiment was rather short (i.e., 20 minutes task duration). This also creates short-lived relationships. Our field experiment design shows the positive effect of transformational leadership styles on crowd worker perceptions. However, there are voices that claim leadership depends on a solid and sustainable personal relationship between the leader and the follower, which develops over time. Testing this idea will require a longer-term experiment curating a crowd of workers for repeated interactions.

Third, we took several design decisions and used only one operationalization of transformational leadership and emotional capabilities in the CA. For transformational leadership, we relied on three transformational leadership styles. However, further refinements of the transformational leadership styles exist in the instrumental or charismatic leadership theory (Antonakis and House 2014). In charismatic leadership, the styles are further subdivided into detailed means of charismatic leaders (e.g., the provision of metaphors or three-part lists) (Antonakis et al. 2022).

Fourth, for the operationalization of emotional capabilities, we relied on the technique of sentiment analysis in the text written by the crowd workers. The operationalization, thereby, is dependent on the emotions given in the text and the capabilities of the sentiment analysis tool in use. Future work may focus on refining the operationalization of emotional capabilities going hand in hand with a deeper investigation of different crowdsourcing tasks since this determines the interaction with the CA. To conclude, future work should seek to better understand and solve the discrepancy between our results regarding the positive subjective performance of the crowd workers and their negative objective performance through the investigation of the limitations.

5.6. Conclusion

Intelligent UAS are capable of providing emotional support to humans due to the presence of emotional intelligence. This chapter investigates the effect of transformational leadership and emotional capabilities of an intelligent UASs represented by a CA on the relationship between the intelligent UASs and the crowd worker (i.e., LMX). This work measures the objective performance and the subjective experience of effectiveness and enjoyment of the crowd workers. The results show a positive effect of intelligent UASs with transformational leadership capabilities on LMX and a positive stimulation of perceived effectiveness and task enjoyment by the crowd workers. The objective performance in terms of idea quality and elaboration, in turn, was worse in the case of conversational crowdsourcing with transformational leadership and emotional capabilities. Further, the emotional capabilities reduced the positive effect of transformational leadership capabilities on LMX.

With the results, this work contributes to a better understanding of the interaction between users and intelligent UASs. The findings shed light on the contrast of positive perceptions of crowd workers in effectiveness and enjoyment and mitigated objective performance results. Improving

crowd workers' perceptions of their work environment is highly beneficial. The degradation of objective performance is an obstacle that has to be overcome. Future research, therefore, should investigate drivers, characteristics, and potential mechanisms of transformational leadership in intelligent UASs to retain the positive effect on crowd worker perceptions and to reverse the objective performance results. By achieving such an ambitious goal, intelligent UASs could further contribute to a better work-life for a new and increasing digital workforce of the 21st century, the crowd workers.

6. Anticipating User Assistance Systems to Improve Collective Decision-Making

Anthropomorphism describes the attribution of human-like physical or non-physical features, behavior, emotions, characteristics, and attributes to a non-human (Epley et al. 2007). The use of anthropomorphism helps human beings to better explain unknown circumstances during the interaction with non-human entities like information and communication technologies and, thus, to better understand and predict their behavior (Hegel et al. 2008). This approach is supported by the familiarity thesis, which claims that human beings explain new stimuli through a mental model of already familiar stimuli (Guthrie 1997).

The human tendency to humanize socio-technical systems can be used in the development of anthropomorphic IS to reduce emotional distance to the IS or an UAS enclosed in an IS and to create a natural connection between human beings and socio-technical systems or its components (Epley et al. 2007; Pfeuffer et al. 2019). In particular, new technologies make it possible to implement increasingly human-like features that further increase familiarization with IS. Improvements in the field of affective computing, for example, allow to better measure the emotions of individuals during the interaction with IS (Calvo et al. 2015; Hudlicka 2003; Tao and Tan 2005; Wu and Liang 2011). By processing this information, IS can increase their emotional intelligence and react more human-like. At the same time, advances in artificial cognitive intelligence make it possible to solve more complex problems and to perform more human-like activities in group collaboration. Through this, IS can provide context-related and semantically correct responses for increasingly complex problems and thereby increase the perception of human-like cognitive processes. Besides increasing cognitive and emotional intelligence, contemporary and avant-garde interface design including embodiment, avatars, speech and natural language processing further contribute to the perception of human-likeness and anthropomorphism.

By enhancing IS with such complex anthropomorphic cues, it is possible to develop anticipating UASs that adapt to the current context and the preferences of their users as well as predict future needs and actions. Anticipating UASs can support users in fulfilling a task by not only offering advice on a topic but also referring to the user's current activities and environmental conditions in order to provide context-related recommendations and advance interaction between users and anticipating UASs (Maedche et al. 2016). Anticipating UASs also incorporate previous experiences to predict user reactions and shape the interaction accordingly (Maedche et al.

2016). Based on these characteristics and technological progress, it is likely that groups of both humans and anticipating UASs will increasingly interact in a collectively intelligent way (Gimpel 2015). In such collectively intelligent group decision-making settings, anticipating UASs do not only take the role of merely providing tools for humans to communicate and collaborate more effectively. Rather anticipating UASs increasingly take on the role of intelligent social actors collaborating with humans (Gimpel 2015). Currently available examples are customer service assistants (Gnewuch et al. 2017; Lee and Choi 2017) and virtual assistants (Luger and Sellen 2016). In particular, virtual assistants, such as Amazon's Alexa, Google's Now, Apples' Siri and Microsoft's Cortana, are becoming more and more advanced and positively received (Maedche et al. 2016). Nowadays, collaborative human-UAS decision-making settings typically involve only one human and one anticipating UAS. These (very small) groups will become larger involving multiple anticipating UASs, multiple humans, or both. Thus, the development of the theoretical model is not limited to dyadic interaction but to group collaboration involving at least one human and at least one anticipating UAS in general. For simplicity, the proposed experiment design focuses on one-on-one interaction in a dyad. Future empirical work should go beyond this and add evidence for larger groups.

The influence of an anticipating UAS on humans depends on the amount of realism (Blascovich 2002). Although the use of anthropomorphic cues has the goal of establishing a positive connection between humans and anticipating UASs, such as increased trust (Aldiri et al. 2008), higher acceptance (Nass et al. 1995), and more entertainment (van Mulken et al. 1998), negative emotional responses can also occur if the anticipating UAS has characteristics which are very similar to those of humans. The theory of the "uncanny valley" by Mori (1970) states that an increase in human-like appearance and behavior increases the likability of non-human entities up to a point where likability decreases significantly. The reason for this phenomenon is that the human likeness also increases expectations of the non-human entity, which cannot be fulfilled, leading to a feeling of disappointment and repulsion (Bartneck et al. 2009). However, if the degree of human likeness continues to increase and expectations can be fulfilled, this in turn leads to a new rise in likability (Mori 1970).

To ensure the acceptance and spread of anticipating UASs and thereby create successful assistance relationships, it is necessary to better understand how humans react to anthropomorphic cues and how they affect the collaboration with anticipating UASs (Pfeuffer et al. 2019). Earlier research in the field of IS focused on the technical implementation of anthropomorphic cues, such as designing the appearance and movements of robots (Bartneck

et al. 2009; Duffy 2003; Hammer et al. 2017; Häring et al. 2014; Hegel et al. 2008; Walters et al. 2008) and virtual avatars (Benbasat et al. 2010; Etemad-Sajadi and Ghachem 2015; Gratch et al. 2002; Nowak 2004; Nowak and Biocca 2003; Nowak and Rauh 2005, 2008; von der Pütten et al. 2010). Furthermore, researchers investigated the interaction between UASs and humans, but the results are limited to supporting functions (Krumeich et al. 2015; Maedche et al. 2016; Schacht et al. 2014). Research to date has hardly addressed the impact of anthropomorphic cues of anticipating UASs on the interaction in which anticipating UASs act as intelligent social actors collaborating with human-beings. It is unclear whether the positive effects of the use of anthropomorphic cues also lead to an improvement in decision quality made in the collaboration process. Based on this background, the overarching research question for this chapter is: In the context of group decision-making by collectives of humans and anticipating UASs, what are the effects of anthropomorphism on the quality of the decision, the satisfaction with the decision, and the personal responsibility for the decision?

The remainder of this chapter is structured as follows. Section 6.1 presents the theoretical background. Section 6.2 introduces the theoretical research model. Section 6.3 introduces an exemplary application scenario and sketches the design of a potential laboratory experiment to test the research model.

6.1. Theoretical Background

The present research is based on prior work, especially from the IS, human-computer interaction, (small) group decision-making, and social psychology literature. It builds on knowledge from research areas such as AI, knowledge representation, reasoning techniques, and formal computational modeling (Gnewuch et al. 2017; Maedche et al. 2016; Silverman 1992; Terveen 1995). Additionally, findings from the field of human-computer interaction, such as information presentation, interaction design, and affective computing, are also used (Bickmore et al. 2005b; Duffy 2003; Terveen 1995; van Mulken et al. 1999; Walters et al. 2008). This knowledge is combined with findings from the field of group decision-making, which examines the different forms of collaboration in (human) groups with two or more actors (Cartwright 1968; Kantosalo and Toivonen 2016; Lubart 2005; Silverman 1992). Table 14 defines the core constructs.

Construct	Definition
Anthropomorphic cues	Perceivable characteristics of an IS that cause individuals to anthropomorphize the IS or parts thereof. These characteristics relate to the system's appearance (Bartneck et al. 2009; Walters et al. 2008), interaction (Gnewuch et al. 2017; Kuchenbrandt et al. 2014), and cognitive abilities (Beale and Creed 2009; Eyssel et al. 2010; Picard et al. 2001).
Anthropomorphism	The attribution of human-like physical or non-physical features, behavior, emotions, characteristics and attributes to a non-human (Epley et al. 2007).
Presence	Presence combines copresence and social presence. Copresence is the psychological connection to and with another person. It requires that interactants feel they were able to perceive their interaction partner, and that their interaction partner actively perceived them (Nowak 2001; Nowak and Biocca 2003). Social presence is "the degree of salience of the other person in the interaction and the consequent salience of the interpersonal relationships" (Short et al. 1976).
Rapport	The feeling of connection and harmony with a partner (Huang et al. 2011). It is characterized by trust and mutual affinity (Vecchi et al. 2005) and deepens over time based on positivity, mutual attention, and coordination (Tickle-Degnen and Rosenthal 1990). We consider short-term rapport building instantly in a short interaction (Cassell et al. 2007).
Group cohesion	The degree to which members of a group are drawn toward one another and have a desire to remain a part of the group (Cartwright 1968). This involves two dimensions: task cohesion and social cohesion (Carron 1982; Carron et al. 1985; Festinger et al. 1950; Mullen and Copper 1994).
Collaboration capabilities	The degree to which an actor like an anticipating UAS can contribute to the operational execution of the collaboration. This includes skills such as context-sensitive communication with humans, agreeing on the shared goal, the provision of knowledge from the problem domain and adaptation based on the previous collaboration history (Duffy 2003; Terveen 1995).
Decision quality	The level of alignment of a decision with the decision a "homo oeconomicus" would have made based on economic theory.
Satisfaction with the decision	The degree of satisfaction the individual experiences with respect to the decision immediately after decision-making.
Outcome	The actual result following a decision based on a risk-related random variable.
Personal responsibility	The degree to which an individual attributes the outcome of a decision to his or her effort and abilities (Campbell and Sedikides 1999).

Table 14: Definition of Core Constructs

6.2. Hypothesis Development

Anticipating UASs can possess various characteristics that lead to humans perceiving them as being more human-like. These anthropomorphic cues can be classified into three categories. First, the category appearance includes visual characteristics such as physical form and visual behavior such as gestures and facial expressions. Previous research has shown that the appearance has a distinct positive influence on the degree of anthropomorphism (Bartneck et al. 2009; Duffy 2003; MacDorman 2006; Nowak and Rauh 2005, 2008; van Mulken et al. 1999; Walters et al. 2008). Second, the category interaction includes different interaction channels between humans and anticipating UASs that allow for more natural interaction, such as semantic text analysis, speech recognition, and speech synthesis. Previous research has shown that the more naturally the interaction, the more it is anthropomorphized (Eyssel et al. 2012; Gnewuch et al. 2017). Third, cognitive abilities contain functionalities that enable the anticipating UAS to process information intelligently, using cognitive intelligence and emotional intelligence. Previous research has demonstrated that the use of more intelligent functionalities leads to an increase in the level of anthropomorphism (Beale and Creed 2009; Eyssel et al. 2010; Gnewuch et al. 2017; Picard et al. 2001). However, increasing the amount and intensity of anthropomorphic cues can also have a negative effect on the perception of human likeness, as described in the uncanny valley theory (Mori 1970). Intensive use of anthropomorphic cues leads to higher expectations until these can no longer be fulfilled, which leads to a reduction in the perceived human likeness (Bartneck et al. 2009). As soon as the expectations are fulfilled, the perceived human likeness increases again (Mori 1970).

Hypothesis 6.1.1: The amount and intensity of anthropomorphic cues with respect to appearance, interaction, and cognitive abilities of the anticipating UAS influence the level of anthropomorphism. The relation is non-monotonic as described by the uncanny valley.

Presence relates to the perception and salience of the interaction partner that results in an interpersonal psychological relationship. The concept originally referred to the presence of other human actors. The “computers as social actors” paradigm (Nass et al. 1999) posits that humans apply social heuristics to interactions with computers. Thus, human users can perceive the copresence and social presence of computer agents (Nowak and Biocca 2003). Bee et al. (2010), for example, found that in the interaction of a human with an embodied CA, interactive gaze (i.e., a greater amount of behavioral realism) was associated with a higher level of perceived presence.

Hypothesis 6.2.1: The stronger the anthropomorphism of a human user towards an anticipating UAS, the higher the presence of the anticipating UAS as perceived by that human user.

In interpersonal relationships, rapport is created through behaviors signaling positive emotions (e.g., head nods or smiles), mutual attention (e.g., mutual gaze), and coordination (e.g., postural mimicry, synchronized movements) (Huang et al. 2011; Tickle-Degnen and Rosenthal 1990). In settings with anticipating UASs exhibiting anthropomorphic cues, proper verbal and nonverbal behaviors of virtual agents have been shown to create rapport during interactions with human users (Huang et al. 2011). Likewise, the aforementioned study on the interaction of a human with an embodied CA, found that interactive gaze caused a higher level of rapport (Bee et al. 2010).

Hypothesis 6.1.3: The stronger the anthropomorphism of a human user towards an anticipating UAS, the higher the user's level of rapport with the anticipating UAS.

Group cohesion consists of task cohesion and social cohesion. Task cohesion describes the degree of consensus of a group to achieve a shared goal (MacCoun et al. 2006). We assume that the higher the presence of an anticipating UAS, that is the stronger and more salient the psychological connection between human user and anticipating UAS, the easier it is for human users to perceive agreement on a shared goal with the anticipating UAS. Social cohesion describes the emotional relationship between group members, including aspects such as liking, closeness, and friendship (MacCoun et al. 2006). We assume that copresence and social presence contribute to the emergence of social cohesion among human users and anticipating UASs.

Hypothesis 6.1.4: The higher the presence of the anticipating UAS as perceived by that human user, the higher the group cohesion perceived by that human user.

Previous research has shown that a higher degree of anthropomorphism leads to a more positive perception of anthropomorphic systems on a personal level, for example, through higher acceptance, more liking, and more trust (Nass et al. 1999; van Mulken et al. 1998; Wexelblat 1998). We assume that a human user's feeling of connection and harmony with an anticipating UAS (i.e., rapport) will increase the desire to continue working with the anticipating UAS (i.e., group cohesion). This is in line with prior research that demonstrated that higher rapport with anthropomorphic CAs goes along with higher social attraction to the character (Bee et al. 2010).

Hypothesis 6.1.5: The higher the user's level of rapport with the anticipating UAS, the higher the group cohesion perceived by that human user.

Group cohesion has frequently been identified as an explanation for group performance (Ahronson and Cameron 2007; Boyle 2003; Carron et al. 2002; Costello 2004; Ensley and Pearson 2005; Klein and Mulvey 1995; Senécal et al. 2008). For example, the perception that the collaboration partners work together to achieve a shared goal leads to a higher motivation of the collaboration partners, to a more productive working atmosphere, and therefore to a higher quality of results (Legler and Reischl 2003; Mullen and Copper 1994).

Hypothesis 6.1.6: The higher group cohesion perceived by the human user, the higher the quality of the decision.

Furthermore, we argue that the higher motivation and more productive atmosphere give people the feeling that working with the anticipating UAS was productive and that an effective decision could be made through collaboration. This perception then has a positive effect on their immediate satisfaction with the decision.

Hypothesis 6.1.7: The higher the group cohesion perceived by the human user, the higher the user's satisfaction with the decision.

To create a successful collaboration between anticipating UASs and humans, it is necessary that humans perceive anticipating UASs as viable collaboration partners capable of supporting the accomplishment of the shared goal. Thus, the perceived collaboration capabilities of anticipating UASs must convince human users of the possibility of a collaboration so that the anticipating UASs are not perceived as an obstacle during collaboration. We assume that the level of anthropomorphism has an influence on the perception of the collaboration capability. Previous research has shown that the display of emotions by virtual agents influenced people's decision-making processes: People cooperated more with virtual agents showing cooperative facial displays than competitive facial displays (de Melo et al. 2015). Likewise, the use of anthropomorphic cues during communication has a positive effect on the perception of knowledgeable-ness of a technical system, the ability to engage with users (Cassell and Bickmore 2003), and credibility (Nowak and Rauh 2008).

Hypothesis 6.1.8: The stronger the anthropomorphism of a human user towards the anticipating UAS, the higher the anticipating UAS's collaboration capabilities perceived by that human user.

Collaboration capabilities are closely related to decision quality. Low perceived collaboration capabilities lead to the anticipating UAS being perceived by humans as a disturbance to collaboration, which makes working together more difficult. This creates a less productive atmosphere, which has a negative influence on the quality of decision-making. High perceived collaboration capabilities enable humans and anticipating UASs to collaborate without disturbance, allowing them to combine their unique abilities (Terveen 1995) to achieve higher decision quality.

Hypothesis 6.1.9: The higher the anticipating UAS's collaboration capabilities perceived by the human user, the higher the quality of the decision.

Based on the more productive collaboration and the combination of unique abilities made possible by a higher degree of perceived collaborative capabilities, we assume that this creates the feeling within humans that they have made better use of the collaboration and worked together more effectively with the anticipating UAS. Thus, the perceived improvement of the use of the collaboration in turn leads to greater satisfaction with the decision.

Hypothesis 6.1.10: The higher the anticipating UAS's collaboration capabilities perceived by the human user, the higher the user's satisfaction with the decision.

Decisions are often subject to risk, which means that the outcome is not only determined by the collaboration process but is also influenced by stochastic external factors. Thus, even a productive collaboration process can lead to a negative outcome, which the collaboration partners are (in part) responsible for. However, previous research shows that an individual's personal responsibility is reduced by the presence of a group, as personal responsibility diffuses within the group (Darley and Latané 1968). Since a higher degree of anthropomorphism leads to a higher perceived presence (Bee et al. 2010; Nowak 2004), it can be concluded that the feeling of being part of a group also increases. Thus, a higher level of anthropomorphism leads to a greater transfer of personal responsibility to the anticipating UAS as a group member, thereby reducing the individual's personal responsibility. Additionally, the outcome after the decision moderates the negative influence of anthropomorphisms on personal responsibility. Earlier research in psychology shows that people tend to attribute negative outcomes to external factors such as other group members, whereby positive outcomes are attributed to personal effort and abilities (Campbell and Sedikides 1999). Thus, a negative outcome increases the negative influence of anthropomorphism on personal responsibility. In contrast, a positive

outcome leads to an increase in personal responsibility and thus counteracts the negative influence of anthropomorphism.

Hypothesis 6.1.11: The stronger the anthropomorphism of a human user towards the anticipating UAS in group decision-making, the lower the level of personal responsibility that user feels once the long-term consequences of the decision become apparent and turn out badly or the higher the level of personal responsibility when the outcome is favorable.

The effects hypothesized here might be moderated by the decision-making task and context as well as by individual characteristics of the human users. Likewise, we assume that emotions play a role in mediating and moderating the hypothesized effects. Three approaches to model emotions can be found in the literature: a categorical approach, which models emotions as distinct categories, such as joy, anger, surprise, fear, or sadness; a dimensional approach, which characterizes emotions in terms of several continuous dimensions, such as arousal, valence, and dominance; and an appraisal-based approach, which describes emotions as valued reactions to emotion-eliciting stimuli. These factors are listed as “control variables” in Figure 20. They should be accounted for in any further elaboration of the research model and in any empirical test.

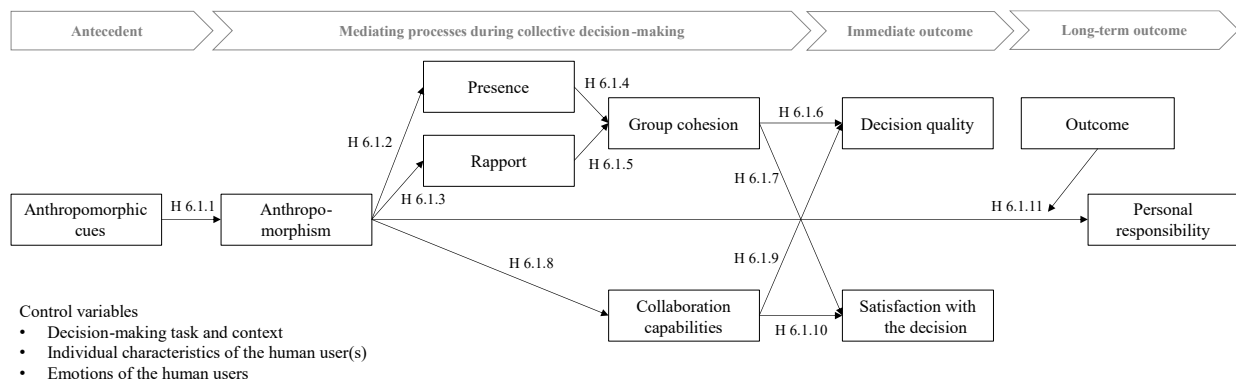


Figure 20: Research Model

6.3. Potential Experiment Design

Exemplary Application Scenario

Consider a financial decision-making scenario where an individual private person aims to make a financial decision like investing for short-term wealth accumulation or investing for long-term retirement provisions. There are (at least) four stylized settings in how the individual can come to a decision

1. The individual might decide on her or his own based on private preferences and knowledge as well as information provided by traditional media like newspapers or online services of his or her bank or broker.

2. The individual might seek the services of a human financial advisor and might in an intense personal discussion come to a decision. The authority to make the decision remains with the individual. However, it is a collective group process among human actors leading to this decision.

3. The individual might use the service of a virtual financial advisor. Like in setting 2, this is a group process leading to an individual decision. In contrast to setting 2, the advisor is not human but a computer actor. Nowadays, first “robo-advisors” are provided by financial services firms (Jung et al. 2018). These contemporary robo-advisors are non-embodied UASs (Maedche et al. 2016) exposing financial optimization models towards laymen.

4. The final setting extends setting 3 with avantgarde technologies: the robo-advisor becomes increasingly intelligent in financial decision-making, increasingly emotionally intelligent, and increasingly human-like in interaction via speech, gestures, and body language of avatars. Thereby, anthropomorphism likely increases and setting 4 gradually moves closer to setting 2.

An underlying assumption of the present research is that technological progress will more and more facilitate setting 4, that providers will offer such services and consumers will start using them. In this exemplary scenario, the abstract research question from this chapter translates to questions like: How satisfied and confident are customers with the investment decision at the time of decision-making? How do they feel about the decision after some time when they experience the effect of their decision, for example, after a crash of stock markets? From a rational financial portfolio investment standpoint, what is the objective quality of the investment decision?




Description of a Potential Experiment Design

For a first empirical test of our hypotheses, this chapter presents a laboratory experiment in which human participants are confronted with a complex decision-making problem under risk. The experiment starts with “very small group decision-making” involving groups of two participants: one human participant, and one anticipating UAS. This extreme case of group decision-making falls within the larger context of collective human-machine decision-making as described above. Having only dyadic interactions increases experimental control as

compared to larger groups. Further research should then gradually increase the number of both humans and anticipating UASs in the groups.

The experiment will use a financial decision-making scenario. Specifically, a scenario of buying or selling risky financial assets in the short-term appears appropriate as it is easier to reconstruct in a laboratory environment than long-term investments. The financial investment domain has the advantage of providing a rather broad body of theoretical research to assess objective decision quality, a rather structured decision setting that is comparatively easy to model and communicate, and a broad literature on experimental finance that allows to transfer elements of the experiment design and procedures to the present research. Classical experiment designs on the disposition effect in financial investments (i.e., investors' propensity to realize past gains more than past losses) like the experiment by Weber and Camerer (1998) might, for example, be adapted to a human-computer group decision-making setting. This allows to draw from a substantial body of experimental finance research on decision quality and responsibility (Fogel and Berry 2006). To incentivize the participants to optimize their decisions, they are paid after the experiment based on their investment performance.

Three experimental treatments are designed in a between-participant design. In each of these treatments, an individual participant collaborating with an anticipating UAS makes decisions on buying or selling risky financial assets. The intensity of anthropomorphic cues is the treatment variable. To compare the experimental groups in terms of decision quality, it is important that the anticipating UASs in different treatments each have identical cognitive abilities with respect to financial decision-making, as otherwise, it is not possible to distinguish whether a potential effect can only be attributed to the degree of anthropomorphism or is caused by the improvement of cognitive abilities. Therefore, the experiment focuses on the appearance and visual behavior of the anticipating UAS (facial expression, gestures) and relational cues like self-disclosure (Moon 2000), empathy (Lisetti et al. 2013), humor (Morkes et al. 1999) and meta-talk (Bickmore and Schulman 2012). Table 15 sketches how the anthropomorphic cues might be configured for the three treatments.

Anthropomorphic cues	Low	Medium	High
Exemplary virtual agent design			
Look	Low human-likeness	High human-likeness	High human-likeness
Facial expressions	Static	Static	Dynamic
Gestures	Static	Static	Dynamic
Self-disclosure	No	Medium	Strong
Empathy	No	Medium	Strong
Humor	No	Medium	Strong
Meta-talk	No	Medium	Strong

Note: Source of images: Adobe Fuse CC (Beta)

Table 15: Exemplary Configuration of Anthropomorphic Cues

Figure 21 sketches the experiment flow as a sequence of 16 steps grouped in 6 phases. For efficiency of the experiment design and sufficient observations for reliable inference, it is beneficial to repeat phases 2 to 4 in multiple rounds for each participant. Pre-tests will need to show whether such a multi-round approach is viable or whether confounding from round to round only allows for one round per participant.

All anthropomorphic cues are presented in Phase 2, the group decision-making phase. Depending on a participant’s engagement with the anticipating UAS, the collaboration involves a debate on the decision task, the participant’s preferences and options, mental models to think through the decision task, and the debate of different scenarios. The anticipating UAS cognitive capabilities to perform this collaboration will be the exact same in all treatments.

Anthropomorphic cues have the potential to enhance the collaboration between users and anticipating UASs. Based on previous research from different disciplines, higher decision quality and satisfaction could be achieved. However, an important next step is to verify the theoretical hypotheses presented in this chapter. Thus, this experimental design should serve as a starting point for future research to test the hypotheses and generate further knowledge on the interaction between users and UASs.

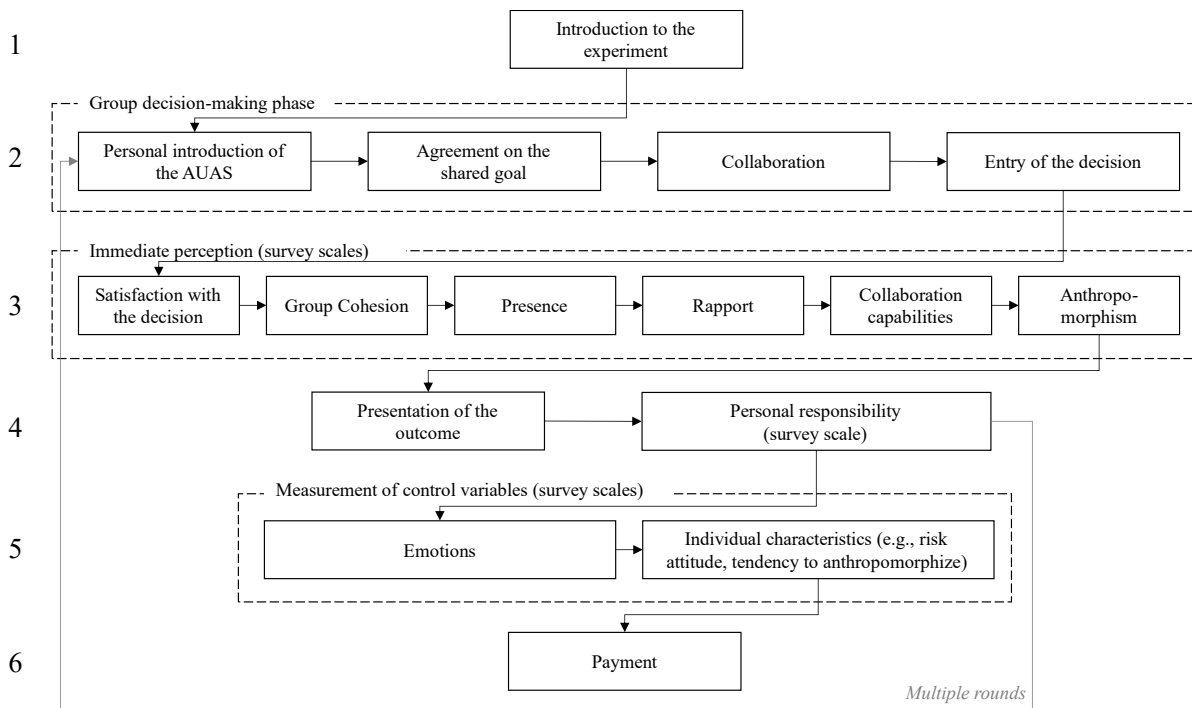


Figure 21: Sequence of Steps to Be Performed by Each Participant

7. General Discussion and Conclusion

7.1. Summary of Results and Meta-Inferences

New technologies enable the development of UASs that assist people in their daily lives and work. UASs are very versatile and thus able to assist in solving different problems of individuals, organizations, and society. However, there is no single solution for the development of UASs. UASs are socio-technical systems characterized by the interaction between users, UASs, and tasks in a specific application context (Maedche et al. 2019). Developers of UASs should ensure a balance between these actors. If these actors are not in balance, for example, if the UASs are too limited or too complex, this results in low acceptance and adaptation (Cesta and D'aloisi 1999; Fuckner et al. 2014; Krogue 2017; McKnight et al. 2002; Zierau et al. 2020). Thus, the potential benefits that UASs can provide would be lost. The IS community has recognized this and is calling for further research on the development of UASs (Dhiman et al. 2022; Maedche et al. 2019). The objective of this dissertation is to better understand the interaction between users, UASs, and tasks for exemplary application contexts and to generate new knowledge that informs future UASs design by presenting prototypes for different types of UASs. For this purpose, four classes of UASs are considered: basic UASs, interactive UASs, intelligent UASs, and anticipating UASs (Maedche et al. 2016).

Chapter 3 presents new insights from the interaction of users with basic UASs to improve the perception of fake news and reduce its spread. People's ability to detect fake news is flawed (Kim and Dennis 2019; Lewandowsky et al. 2012). Therefore, research and practice have presented different interventions but with mixed results (Moravec et al. 2019; Ross et al. 2018). To design effective interventions, it is important to understand the underlying information-processing processes. For example, previous research shows that the order in which information is displayed has an impact on information adaptation (Crozier and Strange 2018; Lewandowsky and van der Linden 2021). In section 3.2, this dissertation conducts an experiment to understand how the order of an intervention provided by a basic UAS affects the perception and adoption of fake news. The basic UAS provides the user with a warning and additional verified information when fake news is encountered. The results confirm that showing an intervention after the fake news story is more beneficial compared to showing the intervention before the fake news story. Showing the intervention before the fake news story leads to a reduction of believability and less adaptation of fake news as well as a higher adaptation of verified

information into the mental model. However, these positive effects decrease over time. After two weeks, the results reveal that there are still fewer fake news stories present in the mental model if the intervention was displayed after the fake news story, but the positive effect of reducing the believability of the fake news stories is lost. Furthermore, after two weeks, there is also no observable effect of intervention order on the adaptation of verified information into the mental model. However, this dissertation reveals that the order in which a basic UAS displays its intervention against fake news has an impact on the processing and perception of fake news.

Basic UASs also have the potential to reduce the spread of fake news. Social media platform providers enable users to participate in the fight against fake news by offering the opportunity to report fake news, but this feature is not sufficiently used. Therefore, this dissertation presents an approach for basic UASs that motivates users to report fake news by using SNs. The results in Section 3.3 show that the use of an injunctive SN leads users to report more fake news. However, no positive effect of descriptive SNs on reporting behavior can be observed. In fact, an overly pronounced descriptive SN reduces users' willingness to report. If injunctive SNs are combined with descriptive SNs, a stronger positive effect on the willingness to report fake news can even be observed than if the types of SNs are used separately. This dissertation shows that a basic UAS can achieve a positive effect on willingness to report fake news by using SNs. Overall, based on a solid theoretical foundation, this thesis demonstrates that basic UASs can support society in the fight against fake news.

Chapter 4 structures and examines the broad body of knowledge on the use of interactive UASs represented by DHRs to achieve behavior change in the SNAP context. SNAP factors are a key trigger for non-communicable diseases (World Health Organization 2020) and thus pose a threat to the health of individuals. Interactive UASs can help users change their behavior and thereby reduce the negative consequences of SNAP factors. Many studies show that the human representation (avatar or embodied agent) of an interactive UAS can facilitate health behavior change (Aljaroodi et al. 2019; Noorbergen et al. 2019). This dissertation conducts a structured literature review to better understand how individual design features of DHRs influence the success of BCIs. The results reveal the most widely used social design features. Further, the work demonstrates how the individual design features influence user perception and thereby trigger BCI functions to achieve behavior change. This relationship indicates that for a successful behavior change, interactive UASs designers should first consider the behavior change mechanism and then select the appropriate social design features.

Chapter 5 reveals new insights into the interaction between intelligent UASs with leadership and emotional capabilities and the user to create a productive and enjoyable work environment. Crowd work is gaining in importance (Burger 2021; Groen et al. 2017), but increasing automation leads to isolation and reduced relatedness to other people. Crowd workers often work alone on a task without contact to colleagues and especially without supervision by a human leader. This lowers the beneficial task-related and emotional stimulation of a human leader. Intelligent UASs with distinct emotional intelligence have the potential to fill this gap. This dissertation conducts an experiment to generate knowledge on the influence of an intelligent UASs with leadership and emotional capabilities on crowd work performance and enjoyment. The results reveal that the use of leadership characteristics improves the relationship between intelligent UASs and users. Contrary to expectations, emotional capabilities reduce this positive effect. Leadership capabilities result in a higher degree of task enjoyment. At the same time, leadership characteristics also increase the users' perceived performance. However, the results reveal that the participants' work performance is reduced by the use of leadership characteristics. Overall, intelligent UASs can be used in crowd working environments to improve the well-being of participants during task processing, but negative effects on performance should be carefully considered.

Chapter 6 explores the collaborative interaction between anticipating UASs and users from a theoretical perspective. Anticipating UASs are the most technologically advanced class of UASs (Maedche et al. 2016), which enables them not only to cooperate but to collaborate with their users. Anthropomorphic design cues are often used to make the interaction intuitive and pleasant for the users. However, this bears the risk of too high expectations, disappointment, and repulsion. Therefore, this dissertation develops a theoretical model to explain the influence of an anticipating UAS with anthropomorphic cues on decision quality, satisfaction with the decision, and responsibility for the outcome. The theoretical results predict that anthropomorphic cues have a positive effect on the respective outcome variables. In addition, this chapter presents an experiment design that will allow future research to empirically validate the theoretical findings.

Theoretical Contribution and Implications for Research

Application of User Assistance Systems to Support Problem Solving

Altogether, this dissertation makes multiple theoretical contributions. It demonstrates theoretically and empirically that UASs can be used to positively influence human behavior

and thereby address challenges faced by individuals, organizations, and society. Moreover, this work presents selected design elements for UASs that influence users' perceptions during interaction, leading to behavioral change.

The research activities provide theoretical and empirical evidence that basic UASs can be infused with SNs to motivate social media users to report fake news. Injunctive SNs motivate users to report fake news even when they are expressed by a basic UAS rather than by a real human being. Although no positive effect of descriptive SNs is observed, this work provides further support that the positive influences of injunctive and descriptive SNs combined have the strongest effect. These new findings add to the portfolio of socio-technical interventions which could reduce the risks associated with fake news. At the same time, this research opens more paths for research. Future research should explore further uses of basic UASs in the context of social media to support users in their daily lives. For example, hate speech or cyberbullying is a problem in social media besides fake news (Wong et al. 2016). Based on the results, it can be assumed that basic UASs with SNs have the potential to activate users to actively oppose or report hate speech as well as cyberbullying and thus assist in its removal. Using other mechanisms is also possible at the same time. In addition to SNs, nudging theory presents different tools (Jesse and Jannach 2021; Weinmann et al. 2016), such as framing or priming, that can be used to extend basic UASs. Future research should identify more design features to develop more effective basic UASs and open up new application contexts.

Based on evidence from previous literature, this dissertation demonstrates that interactive UASs, which are represented as digital humans, can support people in behavior change. The behavior change in this research activity relates to SNAP factors, i.e., less smoking, healthier eating, less alcohol, and more physical activity. This work contributes to the research by summarizing in a structured way the most commonly used social design features and analyzing their effect on user perception and behavior change. These insights should serve as a starting point for future research to improve existing BCIs through the specific use of social design features or to develop new social design features. Further, this work reveals that current research focuses on inventions with positive emotions and neglects negative emotions. Researchers should investigate which design features of interactive UASs influence negative emotions and whether these can trigger a positive behavior change.

This dissertation transfers the theoretical findings of the relationship between a human leader and a human follower to the interaction with an intelligent UAS. This research empirically

confirms that leadership capabilities of an intelligent UAS can lead to an improvement in the relationship between users and UASs. This positive relationship improves users' enjoyment and assessment of their own performance during work and can thus contribute to a positive work environment. In contrast, the presence of leadership characteristics can have a negative effect on actual job performance. These findings contradict previous results. Thus, this work contributes to research by revealing a trade-off between positive user perceptions and negative performance when using an intelligent UAS with leadership characteristics. Future research should examine this trade-off in more detail and attempt to reduce the negative effects on performance. The first step could be to investigate possible causes for the negative impact on performance. Possibly, the additional stimulation from the intelligent UAS caused additional cognitive load or limited the users' ability to perform creative tasks, as it was the case in this experiment. Future research should identify the causes and develop design features that act as countermeasures.

Based on previous research from computer science, social psychology, and group decision-making literature, this dissertation generates insights that in settings where anticipating UASs and users collaborate closely on decision-making, anthropomorphic cues might have a positive effect on collaboration. This work indicates that a higher degree of anthropomorphism improves group cohesion and increases collaboration capabilities. This leads to an improvement in decision quality and the user's satisfaction with the decision. Hence, anthropomorphic cues of an anticipating UAS can contribute to successful collaboration. However, these findings are based solely on previous research. Although a possible experiment design is presented, no empirical validation was conducted. Future research should use the findings to empirically test the hypothesized relationships. One interesting consideration is whether the relationships apply in the bidirectional interaction between a user and an anticipating UAS, but also whether the results transfer to a setting with multiple users or multiple UASs.

This dissertation provides theoretical and empirical evidence that UASs can be used in a variety of ways to support users in different application contexts and motivate them to behave positively. The research activities contribute to the existing knowledge on UASs and open up new application contexts.

Better Understanding of the Interaction Between Users and User Assistance Systems

In addition to this problem-oriented perspective, this dissertation also generates new theoretical knowledge to better understand the interaction between users and UASs. The results on

information processing during the interaction with a basic UAS with fake news intervention demonstrate that not only the type of information presented has an impact on the user's perception but also the order in which the information is presented. Inoculation theory recommends showing the intervention with additional information before the fake news (Azzopardi 2021; Lewandowsky and van der Linden 2021), while the misinformation effect indicates that the intervention should be shown after the fake news (Crozier and Strange 2018). This work shows that interventions with additional information should be shown after the fake news. Conflicting information triggers a state of cognitive dissonance, the resolving of which requires cognitive effort (Moravec et al. 2019). If users see the verified information first, there is no motivation to look more closely at the content and resolve the conflict. If users see the fake news story first, they are more inclined to resolve the conflict so that they do not deliberately adapt the false information. This phenomenon motivates participants to engage with the content and identify the questionable content in the fake news leading to a reduction of credibility as well as a lower adaptation of fake news into the mental model. Although this effect decreases over time, this research activity shows that it is important to understand the mental processes during the interaction between UASs and humans in the different application contexts to enable effective assistance.

This dissertation further contributes to research by analyzing the general mechanisms involved in the interaction between interactive UASs and the user when implementing a BCI in the SNAP domain. This work derives that the design features of interactive UASs with DHRs have an impact on user perception and trigger different psychological constructs. These psychological constructs in turn influence BCI function and thus have an effect on BCI success. To ensure the success of a BCI, it is necessary to understand and incorporate the underlying effects. The results suggest that researchers should choose the intervention function first and then prioritize the psychological constructs that support these intervention functions. The design of the interactive UAS should then be chosen accordingly to address the psychological constructs.

Further, this dissertation presents a theoretical model that attempts to explain the influence of anthropomorphic cues of an anticipating UAS on decision quality, satisfaction, and personal responsibility. Based on theoretical evidence, this work derives that anthropomorphic cues have a positive effect on decision quality and satisfaction. However, the results reveal that anthropomorphic cues lead to lower personal responsibility for the outcome, especially when the outcome of the collaboration is negative. The worse the outcome, the more the user blames the anticipating UAS. The more positive the outcome, the higher the share of success that users

attribute to themselves. This trade-off must be considered when designing anticipating UASs. In addition, this work provides an exemplary implementation for an experiment for future research to verify the theoretical results.

The research activities in this dissertation present new theoretical insights to better understand the interaction between users and UASs. This descriptive knowledge can support future research in the development of UASs, but also points to avenues for future research. First, more research is needed to better understand the effect of existing design elements on user perceptions. In the context of this work, it is shown which design elements are used and which psychological reactions are triggered in the user. Future research should seek to better understand which design features have the most importance for and biggest effect on user perception. Furthermore, new design features should be explored and their usability should be evaluated. For example, the research field of digital nudging provides a variety of approaches that could be used to shape the interaction between users and UASs (Jesse and Jannach 2021; Weinmann et al. 2016). Second, further empirical research is needed to validate the theoretical findings. This includes investigating the relationship between selected design features on perception and outcome. A first direct starting point could be the implementation of the presented experiment design for the investigation of the influence of anthropomorphism of anticipating UASs. An empirical investigation could confirm the derived relationships or uncover new phenomena.

New Knowledge for Future Design of User Assistance Systems

This dissertation further contributes to research by developing new knowledge that informs the future design of UASs. It presents different prototype implementations for two basic UASs as well as one intelligent UAS and validates their empirical effectiveness. Additionally, based on previous research, this work presents a theoretical design for an anthropomorphic anticipating UAS. These ideas can be used by future research to develop new UASs. Three design features that future research should consider are highlighted in the following.

First, this work finds further support that anthropomorphizing of UASs has a positive effect on the interaction and relationship between UASs and users. For example, the research on interactive UASs and anticipating UASs suggests that a higher degree of anthropomorphism leads to more acceptance as well as trust and ultimately has a positive effect on the outcome of the interaction. It is expected that this effect can also be transferred to other classes of UASs and other application contexts. For example, an important aspect of supporting basic UASs in

the fight against fake news is that users trust the information and recommendations provided by UASs. A higher degree of anthropomorphism could potentially lead to a higher initial trust and thus to a higher acceptance as well as effectiveness of the interventions used. For example, interventions could be presented not only as plain text but in a speech bubble of an avatar. Of course, the specific beneficial characteristics of basic UASs should not be neglected during development, i.e., the avatar should not interfere too intensively with the user experience. Nevertheless, this is a promising path for further research.

Second, depending on the application context, user assistance should not be seen as a one-time event but as a continuously repeating process. The findings reveal that the positive effects of UASs can be reduced over time, as, for example, important information is forgotten or motivation is lost without reactivation. The results on basic UASs reveal that the positive effects of fake news interventions are reduced. Additionally, the results show that even if users adopt less fake news into the mental model over time, this does not lead to a reduction in believability. Activating the mental model involves cognitive effort, which users in System-1-thinking avoid. Therefore, users would benefit from continuous assistance that motivates them to activate System-2-thinking and to critically engage with the content of fake news. This assumption is reinforced by the results on interactive UASs, such as virtual coaches and fitness trainers. Previous research shows a positive effect of regular notifications, reminders, and feedback on user behavior (Creed and Beale 2012; Lisetti 2009; Olafsson et al. 2019). However, a positive effect of constant assistance is not certain. It can also happen that too much interaction with UASs is perceived as annoying and intrusive. Transferred to the basic UASs for fake news support, the following scenario might occur. People use social media for pleasure and entertainment (Moravec et al. 2018). Therefore, they want to invest little cognitive resources in this activity (Lewandowsky et al. 2012; Moravec et al. 2018). A basic UAS that constantly motivates users to apply System-2-thinking could hinder these objectives and be perceived as a disruptive burden, reducing trust and willingness to use. Although continuous assistance offers several potential benefits, it is not possible to draw a general conclusion about its usefulness. Researchers developing UASs should always consider the interaction of users, UASs, tasks, and application contexts. Therefore, future research should further investigate and empirically validate the usability of continuous user assistance in different application contexts.

Third, design features can backfire. In other words, the thoughtless use of design features can lead to opposite negative effects. The use of anthropomorphism can help increase trust and acceptance. However, higher levels of anthropomorphism also increase expectations of the

UASs. If these expectations cannot be satisfied, disappointment and rejection may follow. Furthermore, the results reveal that descriptive SNs can also backfire. It can be observed that the stronger a descriptive SN is, the stronger the willingness to report fake news. Above a certain threshold, however, the opposite effect can be observed and the willingness to report decreases again. The added value of reporting fake news is perceived as too low because so many other users have already reported the fake news story. This reaction reduces the motivation to report fake news. Additionally, the research activities on intelligent UASs reveal that the use of leadership characteristics leads to higher enjoyment during task execution. Contrary to expectations, however, it can also be observed that the performance of the users decreases. Possible causes might be that the stimulation by the intelligent UAS with leadership characteristics limits the creativity of the users or leads to additional cognitive load. Thus, this dissertation provides further theoretical and empirical evidence that the thoughtless use of design features can have unexpected negative consequences. It is important that the use of UASs leads to a positive outcome. However, the potential negative consequences should not be neglected. Therefore, further research is needed to identify such negative effects. For example, researchers could further explore the use of constant assistance described in this section, which could potentially trigger negative psychological reactions. Furthermore, future research should better understand which mechanisms lead to these negative effects and how they can be counteracted. For example, it is not yet clear why leadership characteristics lead to a negative effect on performance.

Overall, this dissertation provides different contributions to research. It demonstrates the applicability of UASs in exemplary application contexts, derives new theoretical knowledge, and generates new knowledge that informs future UAS design. This dissertation focuses on the development of descriptive knowledge and allows to better understand the interaction between UASs and users. The knowledge gained can be used to develop new UASs that respect the preferences of the users in the corresponding application context and thus promote the acceptance and adaptation of UASs. At the same time, the knowledge can serve as a basis for future research. In particular, this dissertation should encourage IS researchers to use the descriptive findings presented and transform them into prescriptive knowledge through the use of design science to develop even more beneficial and effective UASs.

Limitations

The results of this dissertation are subject to limitations. This section presents the overarching limitations of this work. For individual limitations of each research activity, readers are referred to the corresponding sections in the individual chapters.

The primary research method in this dissertation is experimental research. This research method has limitations that also apply to this thesis. The generalizability of findings is limited. For each of the experiments, a sample with certain characteristics was chosen that is particularly suitable for the respective application context, e.g., young social media users or experienced crowd workers. Given the grounding of the experiments in theory, it is expected that the results are transferable. Nevertheless, a verification of the results with a larger and more representative sample is needed. The individual operationalizations of the constructs are also specific features of our experiments. For example, specific designs were developed for the UASs and the individual design features. Although this work demonstrates the applicability of these design features, it is unclear whether divergent operationalizations lead to the same results. The experiments were not conducted in real-world environments. The experiments were deliberately conducted in dedicated online environments that were designed to simulate the real world as closely as possible while still limiting the influence of external factors. In addition, participants were motivated monetarily or by vouchers for participating in the experiments. Testing the results in a real-world setting would further improve generalizability. Measurement of latent constructs is subject to limitations. Latent constructs, such as the mental model or the relationship between leader and follower, are not observable and are accessed through indicators, which, however, only provide a partial perspective on the actual construct. In this work, established measurement instruments were used where possible or new measurement instruments were developed based on previous research. A verification of the results with other measurement methods would additionally increase the significance of the results. The duration of the experiments also limits generalizability. UASs are designed to assist users in their everyday life and working environment. Therefore, interaction with a UAS usually does not happen only once, but there are several interaction moments. The experiments in this dissertation primarily look at the initial interaction between users and UASs. Although one of the experiments examines the effects of this interaction after two weeks, the interaction between users and UASs happens only once. How the interaction might evolve over a long period of time as users have repeated contact with the UAS and whether the phenomena described in this work change is unclear. This dissertation establishes the foundation as important relationships

are initially revealed. However, future research should investigate the long-term consistency of the results.

The structured literature review also has limitations for this dissertation. The developed search string was chosen as broad as possible to cover all the relevant literature, but it cannot be ensured that relevant literature was not missed. Also, the inclusion and exclusion criteria ensured that the relevant literature on DHRs was identified. Nevertheless, it is possible that knowledge from other domains, such as robotics or avatars with non-human representation, would have contributed to additional findings. Furthermore, only academic literature was used to guarantee the quality of the results. Gray literature and practice-oriented publications may still contain complementary insights that have been neglected in the research community. Although there is confidence that this dissertation has covered the relevant publications to answer the research question, there is an opportunity for future research to enrich the findings with insights from other domains and practice.

In addition to these methodological limitations, the work is also limited by the examination of exemplary socio-technical systems. This dissertation examines the interaction between users and UASs during task execution in exemplary application contexts. As a result, this work provides important insights for specific groups of UASs in specific application contexts.

However, caution is needed when transferring the results to other groups of UASs or other application contexts. For example, this work shows that basic UASs with SNs can be used in the fight against fake news. However, it remains unclear whether basic UASs with SNs are also suitable to assist in successful BCI in the SNAP domain. Furthermore, it is unclear whether interactive UASs can be used in the fight against fake news or whether interactive design is too invasive to social media users. Thus, research is needed to investigate the applicability of UASs in other application contexts. Nevertheless, this work provides important insights on which future research can build. The following section presents the implications of this dissertation for practice.

Implications for Practice

The research in this dissertation has implications for practitioners who design UASs.

First, this thesis presents concrete approaches to UASs that are capable of improving the lives of individuals. Fake news poses a threat to individuals and society (Allcott and Gentzkow 2017; Shu et al. 2017). The findings on mental processes and the use of social media in a basic UASs

can be used by social media platform providers to implement new interventions against fake news. Injunctive SNs are a simple way to motivate users to report fake news. To motivate users even more, practitioners can additionally use a combination of injunctive SNs and descriptive SNs. At the same time, practitioners should question existing interventions and, if necessary, adjust the order in which information is presented in relation to fake news.

In addition to fake news, an unhealthy lifestyle in terms of SNAP is a big threat to individuals (World Health Organization 2020). Therefore, this dissertation highlights the most commonly used design features of interactive UASs that are used for behavior change towards a healthier way of life. This collection is intended to serve as a toolbox for designers of health and coaching apps to develop new interactive UASs or improve existing ones.

Individuals also encounter challenges in their workplace. The increasing trend towards crowd work leads to reduced relatedness to other people. To counteract this phenomenon, this dissertation presents an approach for intelligent UASs with leadership characteristics and emotional characteristics to establish a relationship with the user. The developers of crowd working platforms, but also the providers of crowd working tasks can use the findings to build a more pleasant environment for crowd workers and remain attractive for crowd workers in the long term. For example, they can implement intelligent UASs such as chatbots with leadership characteristics to improve the enjoyment of the crowd workers. However, leadership characteristics can lead to a reduction in performance. Practitioners should be aware of this trade-off and incorporate it into their decision.

Finally, it can be observed that the capabilities of anticipating UASs continue to increase due to new technological developments, which increases the potential for stronger collaboration between UASs and users (Maedche et al. 2016). This dissertation presents anthropomorphism as an important design feature that can positively influence interaction during shared decision-making. Even though the findings are theoretical and have not yet been empirically validated, it seems reasonable for developers of collaboration tools to equip them with anthropomorphic cues in order to improve decision quality and satisfaction with the decision. However, attention should be paid to the fitting of anthropomorphic cues to the capabilities of the UAS. It should be avoided to create too high expectations, which then cannot be satisfied.

Second, developers of UASs should consider the objectives of the assistance and which information processing processes are involved before implementing design features. For example, if the objective is to activate a specific behavior, developers should first consider

which BCI functions support the targeted behavior. Then, it should be analyzed by which psychological reactions or emotions these BCI functions are triggered. Finally, the corresponding design features should be determined. Although this process was developed in this dissertation specifically in the context of interactive UASs and DHRs, the applicability is not limited to this application context, but could serve developers as a basic design pattern to develop effective user-centric UASs.

Third, developers of UASs need to be aware of the potential negative consequences of design features. For example, design features may have the opposite effect in some circumstances. For example, a too strong descriptive SN can lead to a reduction in willingness to report fake news, or a too high degree of anthropomorphism can result in unfulfillable expectations, leading to disappointment and rejection. Moreover, individual design features can have the desired effect but lead to negative effects in other areas. For example, this work reveals that leadership characteristics of a UAS may lead to more enjoyment as desired, but on the other hand, also result in a lowering of work performance. Developers of UASs should be cautious when implementing design features. They should first understand the effects of the design features and assess whether their use is advantageous.

This dissertation presents different insights for developing and interacting with UASs that practitioners can benefit from. This enables the development of even more effective UASs that support users in everyday life or in their working environment.

7.2. Conclusion

UASs offer potential benefits to individuals, organizations, and society. However, to realize these benefits, the hurdles of adaptation must be overcome by designing UASs that understand and address the needs of users (Maedche et al. 2016). Responding to the call for more theoretical as well as practical-relevant research on UASs (Dhiman et al. 2022; Maedche et al. 2019), the research in this dissertation aims to contribute to IS knowledge on UAS design by better understanding the interaction between UASs and users in different application contexts. New theoretical knowledge is developed based on previous research from different research disciplines. Thereby, it contributes to a deeper theoretical and empirical understanding of the information processing processes and psychological response during the interaction between users and UASs. The knowledge gained is empirically evaluated to explain and predict user behavior. At the same time, new knowledge is generated through the development of

prototypical UASs that informs future UASs design. Thereby, this dissertation provides novel findings across different classes of UASs and across different application contexts. The findings should be used by researchers and practitioners to build user-oriented UASs in order to reduce the barriers of adaptation and realize the potential benefits.

References

- Abdullah, A. S., Gaehde, S., and Bickmore, T. W. 2018. "A Tablet Based Embodied Conversational Agent to Promote Smoking Cessation Among Veterans: A Feasibility Study," *Journal of Epidemiology and Global Health* (8:3-4), pp. 225-230.
- Abualsaud, M., and Smucker, M. D. 2019. "Exposure and Order Effects of Misinformation on Health Search Decisions," in *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, Rome, Italy.
- Acciarini, C., Brunetta, F., and Boccardelli, P. 2021. "Cognitive Biases and Decision-Making Strategies in Times of Change: A Systematic Literature Review," *Management Decision* (59:3), pp. 638-652.
- Adam, M. T. P., Dreyer, S., Gimpel, H., and Olenberger, C. 2022. "Digital Human Representations for Health Behavior Change: A Structured Literature Review," *AIS Transactions on Human-Computer Interaction* (14:3), pp. 314-355.
- Agresti, A., and Kateri, M. 2011. "Categorical Data Analysis," in *International Encyclopedia of Statistical Science*, M. Lovric (ed.), Berlin, Heidelberg: Springer.
- Ahn, S. J. 2015. "Incorporating Immersive Virtual Environments in Health Promotion Campaigns: A Construal Level Theory Approach," *Health Communication* (30:6), pp. 545-556.
- Ahn, S. J., Bailenson, J. N., and Park, D. 2014a. "Short- and Long-Term Effects of Embodied Experiences in Immersive Virtual Environments on Environmental Locus of Control and Behavior," *Computers in Human Behavior* (39), pp. 235-245.
- Ahn, S. J., Fox, J., and Hahm, J. M. 2014b. "Using Virtual Doppelgängers to Increase Personal Relevance of Health Risk Communication," in *Intelligent Virtual Agents*, T. W. Bickmore, S. Marsella and C. Sidner (eds.), Springer, pp. 1-12.
- Ahronson, A., and Cameron, J. E. 2007. "The Nature and Consequences of Group Cohesion in a Military Sample," *Military Psychology* (19:1), pp. 9-25.
- Aisch, G., Huang, J., and Kang, C. 2016. "Dissecting the #PizzaGate Conspiracy Theories," available at <https://www.nytimes.com/interactive/2016/12/10/business/media/pizzagate.html>, accessed on Jun 9 2019.
- Aldiri, K., Hobbs, D., and Qahwaji, R. 2008. "The Human Face of E-Business: Engendering Consumer Initial Trust Through the Use of Images of Sales Personnel on E-Commerce Web Sites," *International Journal of E-Business Research* (4:4), pp. 58-78.

- Aljaroodi, H. M., Adam, M. T. P., Chiong, R., and Teubner, T. 2019. "Avatars and Embodied Agents in Experimental Information Systems Research: A Systematic Review and Conceptual Framework," *Australasian Journal of Information Systems* (23), pp. 1-37.
- Aljaroodi, H. M., Chiong, R., and Adam, M. T. P. 2020. "Exploring the Design of Avatars for Users From Arabian Culture Through a Hybrid Approach of Deductive and Inductive Reasoning," *Computers in Human Behavior* (106), pp. 1-14.
- Allcott, H., and Gentzkow, M. 2017. "Social Media and Fake News in the 2016 Election," *Journal of Economic Perspectives* (31:2), pp. 211-236.
- Andrade, A. D., Idrees, T., Karanam, C., Anam, R., and Ruiz, J. G. 2016. "Effects of an Avatar-Based Anti-Smoking Game on Smoking Cessation Intent," *Studies in Health Technology and Informatics* (220), pp. 15-18.
- André, E. 2021. "Socially Interactive Artificial Intelligence: Past, Present and Future," in *Proceedings of the 2021 International Conference on Multimodal Interaction*, Montréal, Canada, p. 4.
- André, E., Bayer, S., Benke, I., Benlian, A., Cummins, N., Gimpel, H., Hinz, O., Kersting, K., Maedche, A., Muehlhaeuser, M., Riemann, J., Schuller, B. W., and Weber, K. 2019. "Humane Anthropomorphic Agents: The Quest for the Outcome Measure," in *Proceedings of the Pre-ICIS Workshop "Values and Ethics in the Digital Age*, Munich, Germany.
- André, E., Gimpel, H., and Olenberger, C. 2018. "An Investigation of the Effects of Anthropomorphism in Collective Human-Machine Decision-Making," in *Proceedings of the 26th European Conference on Information Systems, Workshops and Tutorials*, Portsmouth, UK.
- Antonakis, J., d'Adda, G., Weber, R. A., and Zehnder, C. 2022. "'Just Words? Just Speeches?'" On the Economic Value of Charismatic Leadership," *Management Science* (68:9), pp. 6355-6381.
- Antonakis, J., and House, R. J. 2014. "Instrumental Leadership: Measurement and Extension of Transformational–Transactional Leadership Theory," *The Leadership Quarterly* (25:4), pp. 746-771.
- Arendt, F., Haim, M., and Beck, J. 2019. "Fake News, Warnings and Perceptual Truth: On the Differential Susceptibility for Fake News Depending on Recipients' Political Orientations," *Publizistik* (64), pp. 181-204.

- Ashkanasy, N. M., and Daus, C. S. 2002. "Emotion in the Workplace: The New Challenge for Managers," *Academy of Management Perspectives* (16:1), pp. 76-86.
- Ayers, M. S., and Reder, L. M. 1998. "A Theoretical Review of the Misinformation Effect: Predictions From an Activation-Based Memory Model," *Psychonomic Bulletin & Review* (5:1), pp. 1-21.
- Azzopardi, L. 2021. "Cognitive Biases in Search," in *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval*, Canberra, Australia, pp. 27-37.
- Bae, S. M., Ha, S. H., and Park, S. C. 2005. "A Web-Based System for Analyzing the Voices of Call Center Customers in the Service Industry," *Expert Systems with Applications* (28:1), pp. 29-41.
- Balmas, M. 2012. "When Fake News Becomes Real," *Communication Research* (41:3), pp. 430-454.
- Bandura, A. 1986. *Social Foundations of Thought and Action: A Social Cognitive Theory*, Hoboken, USA: Prentice-Hall.
- Barron, G., and Yechiam, E. 2002. "Private E-mail Requests and the Diffusion of Responsibility," *Computers in Human Behavior* (18:5), pp. 507-520.
- Barthel, M., Mitchell, A., and Holcomb, J. 2016. "Many Americans Believe Fake News Is Sowing Confusion," available at <http://www.journalism.org/2016/12/15/many-americans-believe-fake-news-is-sowing-confusion>, accessed on Feb 21 2020.
- Bartneck, C., Kulić, D., Croft, E., and Zoghbi, S. 2009. "Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots," *International Journal of Social Robotics* (1), pp. 71-81.
- Baskerville, R., Baiyere, A., Gergor, S., Hevner, A. R., and Rossi, M. 2018. "Design Science Research Contributions: Finding a Balance Between Artifact and Theory," *Journal of the Association for Information Systems* (19:5), pp. 358-376.
- Bass, B. M., and Avolio, B. J. 1993. "Transformational Leadership and Organizational Culture," *Public Administration Quarterly* (117:1), pp. 112-121.
- Bass, B. M., Avolio, B. J., Jung, D. I., and Berson, Y. 2003. "Predicting Unit Performance by Assessing Transformational and Transactional Leadership," *The Journal of Applied Psychology* (88:2), pp. 207-218.
- Bass, B. M., and Steidlmeier, P. 1999. "Ethics, Character, and Authentic Transformational Leadership Behavior," *The Leadership Quarterly* (10:2), pp. 181-217.

- Baumgartner, S. E., Valkenburg, P. M., and Peter, J. 2011. "The Influence of Descriptive and Injunctive Peer Norms on Adolescents' Risky Sexual Online Behavior," *Cyberpsychology, Behavior and Social Networking* (14:12), pp. 753-758.
- Beale, R., and Creed, C. 2009. "Affective Interaction: How Emotional Agents Affect Users," *International Journal of Human-Computer Studies* (67:9), pp. 755-776.
- Bee, N., Wagner, J., André, E., Vogt, T., Charles, F., Pizzi, D., and Cavazza, M. 2010. "Discovering Eye Gaze Behavior During Human-Agent Conversation in an Interactive Storytelling Application," in *International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction*, Beijing, China.
- Behm-Morawitz, E. 2013. "Mirrored Selves: The Influence of Self-Presence in a Virtual World on Health, Appearance, and Well-Being," *Computers in Human Behavior* (29:1), pp. 119-128.
- Behm-Morawitz, E., Lewallen, J., and Choi, G. 2016. "A Second Chance at Health: How a 3D Virtual World Can Improve Health Self-Efficacy for Weight Loss Management Among Adults," *Cyberpsychology, Behavior and Social Networking* (19:2), pp. 74-79.
- Benbasat, I., Dimoka, A., Pavlou, P. A., and Qiu, L. 2010. "Incorporating Social Presence in the Design of the Anthropomorphic Interface of Recommendation Agents: Insights from an fMRI Study," in *Proceedings of the 31th International Conference on Information System*, Saint Louis, USA.
- Benke, I., Gimpel, H., Maedche, A., and Olenberger, C. forthcoming. "Leading the Crowd: The Impact of Transformational Leadership and Emotional Capabilities in Conversational Crowdsourcing," *Under review*.
- Beňuš, Š., Gravano, A., and Hirschberg, J. 2011. "Pragmatic Aspects of Temporal Accommodation in Turn-Taking," *Journal of Pragmatics* (43:12), pp. 3001-3027.
- Benyon, D. 2014. *Designing Interactive Systems: A Comprehensive Guide to HCI, UX and Interaction Design*, Harlow, UK: Pearson Education.
- Bernard, J., Dennis, A. R., Galletta, D., Khan, A., and Webster, J. 2019. "The Tangled Web: Studying Online Fake News," in *Proceedings of the 40th International Conference on Information Systems*, Munich, Germany.
- Bernedo, M., Ferraro, P. J., and Price, M. 2014. "The Persistent Impacts of Norm-Based Messaging and Their Implications for Water Conservation," *Journal of Consumer Policy* (37:3), pp. 437-452.

- Bhattacharyya, S., Banerjee, S., Bose, I., and Kankanhalli, A. 2020. "Temporal Effects of Repeated Recognition and Lack of Recognition on Online Community Contributions," *Journal of Management Information Systems* (37:2), pp. 536-562.
- Bickmore, T. W., Caruso, L., Clough-Gorr, K., and Heeren, T. 2005a. "'It's Just Like You Talk to a Friend': Relational Agents for Older Adults," *Interacting With Computers* (17:6), pp. 711-735.
- Bickmore, T. W., Gruber, A., and Picard, R. W. 2005b. "Establishing the Computer-Patient Working Alliance in Automated Health Behavior Change Interventions," *Patient Education and Counseling* (59:1), pp. 21-30.
- Bickmore, T. W., Mauer, D., Crespo, F., and Brown, T. 2007. "Persuasion, Task Interruption and Health Regimen Adherence," in *Persuasive Technology*, D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, Y. de Kort, W. IJsselsteijn, C. Midden, B. Eggen and B. J. Fogg (eds.), Berlin, Germany: Springer, pp. 1-11.
- Bickmore, T. W., and Schulman, D. 2012. "Empirical Validation of an Accommodation Theory-Based Model of User-Agent Relationship," in *Intelligent Virtual Agents*, D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, Y. Nakano, M. Neff, A. Paiva and M. Walker (eds.), Berlin, Germany: Springer, pp. 390-403.
- Bickmore, T. W., Schulman, D., and Sidner, C. 2013a. "Automated Interventions for Multiple Health Behaviors Using Conversational Agents," *Patient Education and Counseling* (92:2), pp. 142-148.
- Bickmore, T. W., Schulman, D., and Yin, L. 2010. "Maintaining Engagement in Long-term Interventions with Relational Agents," *Applied Artificial Intelligence* (24:6), pp. 648-666.
- Bickmore, T. W., Silliman, R. A., Nelson, K., Cheng, D. M., Winter, M., Henault, L., and Paasche-Orlow, M. K. 2013b. "A Randomized Controlled Trial of an Automated Exercise Coach for Older Adults," *Journal of the American Geriatrics Society* (61:10), pp. 1676-1683.
- Birk, M. V., Atkins, C., Bowey, J. T., and Mandryk, R. L. 2016. "Fostering Intrinsic Motivation through Avatar Identification in Digital Games," in *Proceedings of the 2016*

- CHI Conference on Human Factors in Computing Systems*, San Jose, USA, pp. 2982-2995.
- Birk, M. V., and Mandryk, R. L. 2018. "Combating Attrition in Digital Self-Improvement Programs Using Avatar Customization," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, Montreal, Canada, pp. 1-15.
- Blair, C. A., Foster Thompson, L., and Wuensch, K. L. 2005. "Electronic Helping Behavior: The Virtual Presence of Others Makes a Difference," *Basic and Applied Social Psychology* (27:2), pp. 171-178.
- Blascovich, J. 2002. "A Theoretical Model of Social Influence for Increasing the Utility of Collaborative Virtual Environments," in *Proceedings of the 4th International Conference on Collaborative Virtual Environments*, Bonn, Germany, pp. 25-30.
- Blohm, I., Bretschneider, U., Leimeister, J. M., and Krcmar, H. 2011. "Does Collaboration Among Participants Lead to Better Ideas in It-Based Idea Competitions? An Empirical Investigation," *International Journal of Networking and Virtual Organisations* (9:2), pp. 106-122.
- Bono, J. E., and Ilies, R. 2006. "Charisma, Positive Emotions and Mood Contagion," *The Leadership Quarterly* (17:4), pp. 317-334.
- Bono, J. E., and Judge, T. A. 2004. "Personality and Transformational and Transactional Leadership: A Meta-Analysis," *The Journal of Applied Psychology* (89:5), pp. 901-910.
- Boyle, I. T. 2003. *The Impact of Adventure-Based Training on Team Cohesion and Psychological Skills Development in Elite Sporting Teams*, Irvine, USA: Universal-Publishers.
- Brashier, N. M., Pennycook, G., Berinsky, A. J., and Rand, D. G. 2021. "Timing Matters When Correcting Fake News," *Proceedings of the National Academy of Sciences of the United States of America* (118:5), pp. 1-3.
- Braun, K. A., and Loftus, E. F. 1998. "Advertising's Misinformation Effect," *Applied Cognitive Psychology* (12:6), pp. 569-591.
- Britt, M. A., Rouet, J.-F., Blaum, D., and Millis, K. 2019. "A Reasoned Approach to Dealing With Fake News," *Policy Insights from the Behavioral and Brain Sciences* (6:1), pp. 94-101.
- Bromuri, S., Henkel, A. P., Iren, D., and Urovi, V. 2021. "Using AI to Predict Service Agent Stress From Emotion Patterns in Service Interactions," *Journal of Service Management* (32:4), pp. 581-611.

- Burger, C. 2021. "Beliebtheit von Crowdfunding Nimmt in Unternehmen Zu," available at <https://www.vdi-nachrichten.com/karriere/arbeitsmarkt/beliebtheit-von-crowdfunding-nimmt-in-unternehmen-zu/>, accessed on Dec 4 2022.
- Cafaro, A., Vilhjálmsón, H. H., and Bickmore, T. W. 2016. "First Impressions in Human-Agent Virtual Encounters," *ACM Transactions on Computer-Human Interaction* (23:4), pp. 1-40.
- Calvo, R. A., D'Mello, S., Gratch, J., and Kappas, A. 2015. *The Oxford Handbook of Affective Computing*, Oxford, UK: Oxford University Press.
- Campbell, W. K., and Sedikides, C. 1999. "Self-Threat Magnifies the Self-Serving Bias: A Meta-Analytic Integration," *Review of General Psychology* (3:1), pp. 23-43.
- Caputo, A. 2013. "A Literature Review of Cognitive Biases in Negotiation Processes," *International Journal of Conflict Management* (24:4), pp. 374-398.
- Carman, A. 2018. "Google Assistant Will Proactively Tell You if It Thinks Your Flight Will Be Delayed," available at <https://www.theverge.com/2018/12/17/18144550/google-assistant-flight-delay-notification>, accessed on Dec 16 2022.
- Carron, A. V. 1982. "Cohesiveness in Sport Groups: Interpretations and Considerations," *Journal of Sport Psychology* (4:2), pp. 123-138.
- Carron, A. V., Colman, M. M., Wheeler, J., and Stevens, D. 2002. "Cohesion and Performance in Sport: A Meta Analysis," *Journal of Sport and Exercise Psychology* (24:2), pp. 168-188.
- Carron, A. V., Widmeyer, W. N., and Brawley, L. R. 1985. "The Development of an Instrument to Assess Cohesion in Sport Teams: The Group Environment Questionnaire," *Journal of Sport Psychology* (7:3), pp. 244-266.
- Carter, E., and Knol, C. 2019. "Chatbots - An Organisation's Friend or Foe?" *Research in Hospitality Management* (9:2), pp. 113-116.
- Cartwright, D. 1968. "The Nature of Group Cohesiveness," *Group Dynamics: Research and Theory* (91).
- Cassell, J., and Bickmore, T. W. 2003. "Negotiated Collusion: Modeling Social Language and its Relationship Effects in Intelligent Agents," *User Modeling and User-Adapted Interaction* (13), pp. 89-132.
- Cassell, J., Gill, A. J., and Tepper, P. A. 2007. "Coordination in Conversation and Rapport," in *Proceedings of the Workshop on Embodied Language Processing*, Prague, Czech Republic, pp. 41-50.

- Cavazotte, F., Moreno, V., and Hickmann, M. 2012. "Effects of Leader Intelligence, Personality and Emotional Intelligence on Transformational Leadership and Managerial Performance," *The Leadership Quarterly* (23:3), pp. 443-455.
- Cesta, A., and D'aloisi, D. 1999. "Mixed-Initiative Issues in an Agent-Based Meeting Scheduler," in *Computational Models of Mixed-Initiative Interaction*, S. Haller, A. Kobsa and S. McRoy (eds.), Dordrecht, Netherlands: Springer Netherlands, pp. 229-262.
- Cheng, X., Fu, S., Vreede, T. de, Vreede, G.-J. de, Seeber, I., Maier, R., and Weber, B. 2020. "Idea Convergence Quality in Open Innovation Crowdsourcing: A Cognitive Load Perspective," *Journal of Management Information Systems* (37:2), pp. 349-376.
- Chiang, C.-W., Kasunic, A., and Savage, S. 2018. "Crowd Coach: Peer Coaching for Crowd Workers' Skill Growth," *Proceedings of the ACM on Human-Computer Interaction* (2), pp. 1-17.
- Cialdini, R. B. 2003. "Crafting Normative Messages to Protect the Environment," *Current Directions in Psychological Science* (12:4), pp. 105-109.
- Cialdini, R. B. 2007. "Descriptive Social Norms as Underappreciated Sources of Social Control," *Psychometrika* (72:2), pp. 263-268.
- Cialdini, R. B., Demaine, L. J., Sagarin, B. J., Barrett, D. W., Rhoads, K., and Winter, P. L. 2006. "Managing Social Norms for Persuasive Impact," *Social Influence* (1:1), pp. 3-15.
- Cialdini, R. B., Kallgren, C. A., and Reno, R. R. 1991. "A Focus Theory of Normative Conduct: A Theoretical Refinement and Reevaluation of the Role of Norms in Human Behavior," in *Advances in Experimental Social Psychology*, M. P. Zanna (ed.), Amsterdam, Netherlands: Elsevier, pp. 201-234.
- Cialdini, R. B., Reno, R. R., and Kallgren, C. A. 1990. "A Focus Theory of Normative Conduct: Recycling the Concept of Norms to Reduce Littering in Public Places," *Journal of Personality and Social Psychology* (58:6), pp. 1015-1026.
- Clark, O. J., Grogan, S., Cole, J., and Ray, N. 2019. "A Systematic Review on the Influence of Avatar Appearance on Health-Related Outcomes: A Systematic Review and Narrative Meta-Review," *PsyArXiv*.
- Clarke, J., Chen, H., Du, D., and Hu, Y. J. 2021. "Fake News, Investor Attention, and Market Reaction," *Information Systems Research* (32:1), pp. 35-52.
- Cohen, J. 1969. *Statistical Power Analysis for the Behavioral Sciences*, New York, USA: Academic Press.

- Cohen, J. D., Usher, M., and McClelland, J. L. 1998. "A PDP Approach to Set Size Effects Within the Stroop Task: Reply to Kanne, Balota, Spieler, and Faust," *Psychological Review* (105:1), pp. 188-194.
- Connelly, S., and Ruark, G. 2010. "Leadership Style and Activating Potential Moderators of the Relationships Among Leader Emotional Displays and Outcomes," *The Leadership Quarterly* (21:5), pp. 745-764.
- Costello, J. F. 2004. *Developing Team Cohesion: A Quasi-Field Experiment*, Air Force Institute of Technology.
- Creed, C., and Beale, R. 2012. "User Interactions With an Affective Nutritional Coach," *Interacting With Computers* (24:5), pp. 339-350.
- Croson, R., Handy, F., and Shang, J. 2009. "Keeping Up With the Joneses: The Relationship of Perceived Descriptive Social Norms, Social Information, and Charitable Giving," *Nonprofit Management and Leadership* (19:4), pp. 467-489.
- Crozier, W. E., and Strange, D. 2018. "Correcting the Misinformation Effect," *Applied Cognitive Psychology* (33:4), pp. 585-595.
- Cutler, D. M. 2004. "Behavioral Health Interventions: What Works and Why?" in *Critical Perspectives on Racial and Ethnic Differences in Health in Late Life*, N. B. Anderson, A. A. Bulatao and B. Cohen (eds.), Washington, USA: National Academies Press.
- Dahl, D. W., and Moreau, C. P. 2007. "Thinking Inside the Box: Why Consumers Enjoy Constrained Creative Experiences," *Journal of Marketing Research* (44:3), pp. 357-369.
- Dale, R. 2016. "The Return of the Chatbots," *Natural Language Engineering* (22:5), pp. 811-817.
- Darley, J. M., and Latané, B. 1968. "Bystander Intervention in Emergencies: Diffusion of Responsibility," *Journal of Personality and Social Psychology* (8:4), pp. 377-383.
- Das, S., and Shinde, S. 2022. "WhatsApp India, GoI Join Hands to Leverage Chatbot for Non-COVID Use-Cases," available at https://www.business-standard.com/article/current-affairs/whatsapp-india-goi-join-hands-to-leverage-chatbot-for-non-covid-use-cases-122060300649_1.html, accessed on Dec 11 2022.
- Davis, G. B., and Olson, M. H. 1988. *Management Information Systems: Conceptual Foundations, Structure, and Development*, New York, USA: McGraw-Hill.
- Day, A. L., and Carroll, S. A. 2004. "Using an Ability-Based Measure of Emotional Intelligence to Predict Individual Performance, Group Performance, and Group Citizenship Behaviours," *Personality and Individual Differences* (36:6), pp. 1443-1458.

- de Melo, C. M., Gratch, J., and Carnevale, P. J. 2015. "Humans versus Computers: Impact of Emotion Expressions on People's Decision Making," *IEEE Transactions on Affective Computing* (6:2), pp. 127-136.
- de Rosis, F., Novielli, N., Carofiglio, V., Cavalluzzi, A., and Carolis, B. de. 2006. "User Modeling and Adaptation in Health Promotion Dialogs With an Animated Character," *Journal of Biomedical Informatics* (39:5), pp. 514-531.
- de Vries, R. E., Bakker-Pieper, A., and Oostenveld, W. 2010. "Leadership = Communication? The Relations of Leaders' Communication Styles with Leadership Styles, Knowledge Sharing and Leadership Outcomes," *Journal of Business and Psychology* (25:3), pp. 367-380.
- Dean, D., Hender, J., Rodgers, T., and Santanen, E. 2006. "Identifying Quality, Novel, and Creative Ideas: Constructs and Scales for Idea Evaluation," *Journal of the Association for Information Systems* (7:10), pp. 646-699.
- Deci, E. L., Olafsen, A. H., and Ryan, R. M. 2017. "Self-Determination Theory in Work Organizations: The State of a Science," *Annual Review of Organizational Psychology and Organizational Behavior* (4:1), pp. 19-43.
- Deci, E. L., and Vansteenkiste, M. 2004. "Self-Determination Theory and Basic Need Satisfaction: Understanding Human Development in Positive Psychology," *Ricerche di Psicologia* (27:1), pp. 23-40.
- Della Vedova, M. L., Tacchini, E., Moret, S., Ballarin, G., DiPierro, M., and Alfaro, L. de. 2018. "Automatic Online Fake News Detection Combining Content and Social Signals," in *22nd Conference of Open Innovations Association*, Jyvaskyla, Finland, pp. 272-279.
- Demarque, C., Charalambides, L., Hilton, D. J., and Waroquier, L. 2015. "Nudging Sustainable Consumption: The Use of Descriptive Norms to Promote a Minority Behavior in a Realistic Online Shopping Environment," *Journal of Environmental Psychology* (43), pp. 166-174.
- Deng, B., and Chau, M. 2021. "The Effect of the Expressed Anger and Sadness on Online News Believability," *Journal of Management Information Systems* (38:4), pp. 959-988.
- Deng, X. N., Joshi, K. D., and Galliers, R. D. 2016. "The Duality of Empowerment and Marginalization in Microtask Crowdsourcing," *Management Information Systems Quarterly* (40:2), pp. 279-302.
- Dennis, A. R. 1996. "Information Exchange and Use in Small Group Decision Making," *Small Group Research* (27:4), pp. 532-550.

- Dennis, A. R., and Valacich, J. S. 2001. "Conducting Experimental Research in Information Systems," *Communications of the Association for Information Systems* (7), pp. 1-41.
- Dennis, A. R., Valacich, J. S., Connolly, T., and Wynne, B. E. 1996. "Process Structuring in Electronic Brainstorming," *Information Systems Research* (7:2), pp. 268-277.
- Dennis, A. R., Yuan, L., Feng, X., Webb, E., and Hsieh, C. J. 2020. "Digital Nudging: Numeric and Semantic Priming in E-Commerce," *Journal of Management Information Systems* (37:1), pp. 39-65.
- Deterding, S., Dixon, D., Khaled, R., and Nacke, L. 2011. "From Game Design Elements to Gamefulness," in *Proceedings of the 15th International Academic MindTrek Conference on Envisioning Future Media Environments*, Tampere, Finland, pp. 9-15.
- Deutsch, M., and Gerard, H. B. 1955. "A Study of Normative and Informational Social Influences Upon Individual Judgement," *The Journal of Abnormal and Social Psychology* (51:3), pp. 629-636.
- Dhiman, H., Wächter, C., Fellmann, M., and Röcker, C. 2022. "Intelligent Assistants," *Business & Information Systems Engineering* (64:5), pp. 645-665.
- DiFranzo, D., Taylor, S. H., Kazerooni, F., Wherry, O. D., and Bazarova, N. N. 2018. "Upstanding by Design: Bystander Intervention in Cyberbullying," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, Montreal, Canada, pp. 1-12.
- Dissanayake, I., and Sridhar, N. 2021. "The Impact of Helping Others in Coopetitive Crowdsourcing Communities," *Journal of the Association for Information Systems* (22:1), pp. 67-101.
- Duan, C., and Hill, C. E. 1996. "The Current State of Empathy Research," *Journal of Counseling Psychology* (43:3), pp. 261-274.
- Duffy, B. R. 2003. "Anthropomorphism and the Social Robot," *Robotics and Autonomous Systems* (42:3-4), pp. 177-190.
- Dulebohn, J. H., Bommer, W. H., Liden, R. C., Brouer, R. L., and Ferris, G. R. 2012. "A Meta-Analysis of Antecedents and Consequences of Leader-Member Exchange," *Journal of Management* (38:6), pp. 1715-1759.
- Dumdum, U. R., Lowe, K. B., and Avolio, B. J. 2002. "A Meta-analysis of Transformational and Transactional Leadership Correlates of Effectiveness and Satisfaction: An Update and Extension," in *Transformational and Charismatic Leadership: The Road Ahead*, Oxford, UK: Elsevier Science, pp. 35-66.

- Durward, D., Blohm, I., and Leimeister, J. M. 2016. "Crowd Work," *Business & Information Systems Engineering* (58:4), pp. 281-286.
- Durward, D., Blohm, I., and Leimeister, J. M. 2020. "The Nature of Crowd Work and its Effects on Individuals' Work Perception," *Journal of Management Information Systems* (37:1), pp. 66-95.
- Ensley, M. D., and Pearson, A. W. 2005. "An Exploratory Comparison of the Behavioral Dynamics of Top Management Teams in Family and Nonfamily New Ventures: Cohesion, Conflict, Potency, and Consensus," *Entrepreneurship Theory and Practice* (29:3), pp. 267-284.
- Epley, N., Waytz, A., and Cacioppo, J. T. 2007. "On Seeing Human: A Three-Factor Theory of Anthropomorphism," *Psychological Review* (114:4), pp. 864-886.
- Esterwood, C., and Robert, L. P. 2020. "Personality in Healthcare Human Robot Interaction (H-HRI): A Literature Review and Brief Critique," in *Proceedings of the 8th International Conference on Human-Agent Interaction*, Virtual Event, pp. 87-95.
- Etemad-Sajadi, R., and Ghachem, L. 2015. "The Impact of Hedonic and Utilitarian Value of Online Avatars on E-service Quality," *Computers in Human Behavior* (52), pp. 81-86.
- Eyssel, F., Hegel, F., Horstmann, G., and Wagner, C. 2010. "Anthropomorphic Inferences From Emotional Nonverbal Cues: A Case Study," in *19th International Symposium in Robot and Human Interactive Communication*, Viareggio, Italy, pp. 646-651.
- Eyssel, F., Kuchenbrandt, D., Bobinger, S., Rüter, L. de, and Hegel, F. 2012. "'If You Sound Like Me, You Must Be More Human': On the Interplay of Robot and User Features on Human-Robot Acceptance and Anthropomorphism," in *Proceedings of the 7th Annual ACM/IEEE International Conference on Human-Robot Interaction*, Boston, USA, pp. 125-126.
- Facebook. 2016. "Facebook Reports Fourth Quarter and Full Year 2016 Results," Facebook.
- Facebook. 2019. "Company Info," available at <https://newsroom.fb.com/company-info>, accessed on Jan 3 2019.
- Fang, B., Zheng, Z., Ye, Q., and Goes, P. B. 2019. "Social Influence and Monetization of Freemium Social Games," *Journal of Management Information Systems* (36:3), pp. 730-754.
- Feine, J., Gnewuch, U., Morana, S., and Maedche, A. 2019. "A Taxonomy of Social Cues for Conversational Agents," *International Journal of Human-Computer Studies* (132), pp. 138-161.

- Felfe, J., and Schyns, B. 2006. "Personality and the Perception of Transformational Leadership: The Impact of Extraversion, Neuroticism, Personal Need for Structure, and Occupational Self-Efficacy," *Journal of Applied Social Psychology* (36:3), pp. 708-739.
- Ferrara, E., Varol, O., Davis, C., Menczer, F., and Flammini, A. 2016. "The Rise of Social Bots," *Communications of the ACM* (59:7), pp. 96-104.
- Ferris, G. R., Liden, R. C., Munyon, T. P., Summers, J. K., Basik, K. J., and Buckley, M. R. 2009. "Relationships at Work: Toward a Multidimensional Conceptualization of Dyadic Work Relationships," *Journal of Management* (35:6), pp. 1379-1403.
- Fest, S., Kvaløy, O., Nieken, P., and Schöttner, A. 2021. "How (Not) to Motivate Online Workers: Two Controlled Field Experiments on Leadership in the Gig Economy," *The Leadership Quarterly* (32:6), p. 101514.
- Festinger, L. 1957. *A Theory of Cognitive Dissonance*, Stanford, USA: Stanford University Press.
- Festinger, L., Schachter, S., and Back, K. 1950. *Social Pressures in Informal Groups: A Study of Human Factors in Housing*, Stanford, USA: Stanford University Press.
- Fischer, P., Krueger, J. I., Greitemeyer, T., Vogrincic, C., Kastenmüller, A., Frey, D., Heene, M., Wicher, M., and Kainbacher, M. 2011. "The Bystander-Effect: A Meta-Analytic Review on Bystander Intervention in Dangerous and Non-dangerous Emergencies," *Psychological Bulletin* (137:4), pp. 517-537.
- Fischer, T., Dietz, J., and Antonakis, J. 2017. "Leadership Process Models: A Review and Synthesis," *Journal of Management* (43:6), pp. 1726-1753.
- Floyd, R. W. 1986. "Toward Interactive Design of Correct Programs," in *Readings in Artificial Intelligence and Software Engineering*, C. Rich and R. C. Waters (eds.), Amsterdam, Netherlands: Elsevier, pp. 331-334.
- Flynn, D. J., Nyhan, B., and Reifler, J. 2017. "The Nature and Origins of Misperceptions: Understanding False and Unsupported Beliefs About Politics," *Advances in Political Psychology* (38:2), pp. 127-150.
- Fogel, S. O. C., and Berry, T. 2006. "The Disposition Effect and Individual Investor Decisions: The Roles of Regret and Counterfactual Alternatives," *Journal of Behavioral Finance* (7:2), pp. 107-116.
- Fogg, B. J. 2003. "Computers As Persuasive Social Actors," in *Persuasive Technology*, B. J. Fogg (ed.), Morgan Kaufmann, pp. 89-120.

- Følstad, A., and Skjuve, M. 2019. "Chatbots for Customer Service," in *Proceedings of the 1st International Conference on Conversational User Interfaces*, Dublin, Ireland, pp. 1-9.
- Fornell, C., and Larcker, D. F. 1981. "Evaluating Structural Equation Models with Unobservable Variables and Measurement Error," *Journal of Marketing Research* (18:1), pp. 39-50.
- Fox, J., and Bailenson, J. N. 2009. "Virtual Self-Modeling: The Effects of Vicarious Reinforcement and Identification on Exercise Behaviors," *Media Psychology* (12:1), pp. 1-25.
- Fox, J., Bailenson, J. N., and Binney, J. 2009. "Virtual Experiences, Physical Behaviors: The Effect of Presence on Imitation of an Eating Avatar," *Presence: Teleoperators and Virtual Environments* (18:4), pp. 294-303.
- Frederick, S. 2005. "Cognitive Reflection and Decision Making," *Journal of Economic Perspectives* (19:4), pp. 25-42.
- Friederichs, S. A. H., Bolman, C., Oenema, A., Guyaux, J., and Lechner, L. 2014. "Motivational Interviewing in a Web-Based Physical Activity Intervention With an Avatar: Randomized Controlled Trial," *Journal of Medical Internet Research* (16:2).
- Friederichs, S. A. H., Oenema, A., Bolman, C., and Lechner, L. 2015. "Long Term Effects of Self-Determination Theory and Motivational Interviewing in a Web-Based Physical Activity Intervention: Randomized Controlled Trial," *The International Journal of Behavioral Nutrition and Physical Activity* (12:1), pp. 1-13.
- Friedrich, T., Schlauderer, S., Weidinger, J., and Raab, M. 2017. "On the Research Paradigms and Research Methods Employed in the BISE Journal - A Ten-Year Update," in *Proceedings der 13. Internationalen Tagung Wirtschaftsinformatik*, St. Gallen, Switzerland, pp. 1111-1125.
- Fuchs, K., Meusbürger, D., Haldimann, M., and Ilic, A. 2019. "NutritionAvatar: Designing a Future-Self Avatar for Promotion of Balanced, Low-Sodium Diet Intention: Framework Design and User Study," in *Proceedings of the 13th Biannual Conference of the Italian SIGCHI Chapter on Designing the Next Interaction*, Padova, Italy, pp. 1-10.
- Fuckner, M., Barthes, J.-P., and Scalabrin, E. E. 2014. "Using a Personal Assistant for Exploiting Service Interfaces," in *Proceedings of the 2014 IEEE 18th International Conference on Computer Supported Cooperative Work in Design*, Hsinchu, Taiwan, pp. 89-94.

- Füller, J., Mühlbacher, H., Matzler, K., and Jawecki, G. 2009. "Consumer Empowerment Through Internet-Based Co-creation," *Journal of Management Information Systems* (26:3), pp. 71-102.
- Gardiner, P. M., McCue, K. D., Negash, L. M., Cheng, T., White, L. F., Yinusa-Nyahkoon, L., Jack, B. W., and Bickmore, T. W. 2017. "Engaging Women With an Embodied Conversational Agent to Deliver Mindfulness and Lifestyle Recommendations: A Feasibility Randomized Control Trial," *Patient Education and Counseling* (100:9), pp. 1720-1729.
- Gardner, H. 2000. *Intelligence Reframed: Multiple Intelligences for the 21st Century*, New York, USA: Basic Books.
- Garrett, R. K., and Poulsen, S. 2019. "Flagging Facebook Falsehoods: Self-Identified Humor Warnings Outperform Fact Checker and Peer Warnings," *Journal of Computer-Mediated Communication* (24:5), pp. 240-258.
- George, A. S., Muralikrishnan, G., Ninan, L. R., Varrier, P. S., and Dhanya, L. K. 2021a. "Survey on the Design and Development of Indian Language Chatbots," in *2021 International Conference on Communication, Control and Information Sciences*, Idukki, India, pp. 1-6.
- George, J., Gerhart, N., and Torres, R. 2021b. "Uncovering the Truth about Fake News: A Research Model Grounded in Multi-Disciplinary Literature," *Journal of Management Information Systems* (38:4), pp. 1067-1094.
- Gerber, A. S., and Rogers, T. 2009. "Descriptive Social Norms and Motivation to Vote: Everybody's Voting and so Should You," *The Journal of Politics* (71:1), pp. 178-191.
- Gerstner, C. R., and Day, D. V. 1997. "Meta-Analytic Review of Leader–Member Exchange Theory: Correlates and Construct Issues," *Journal of Applied Psychology* (82:6), pp. 827-844.
- Gil de Zúñiga, H., Jung, N., and Valenzuela, S. 2012. "Social Media Use for News and Individuals' Social Capital, Civic Engagement and Political Participation," *Journal of Computer-Mediated Communication* (17:3), pp. 319-336.
- Gilovich, T., Griffin, D. W., and Kahneman, D. 2002. *Heuristics and Biases : The Psychology of Intuitive Judgment*, New York, USA: Cambridge University Press.
- Gilson, L. L., Mathieu, J. E., Shalley, C. E., and Ruddy, T. M. 2005. "Creativity and Standardization: Complementary or Conflicting Drivers of Team Effectiveness?" *Academy of Management Journal* (48:3), pp. 521-531.

- Gimpel, H. 2015. "Interview with Thomas W. Malone on "Collective Intelligence, Climate Change, and the Future of Work"," *Business & Information Systems Engineering* (57:4), pp. 275-278.
- Gimpel, H., Heger, S., Kasper, J., and Schäfer, R. 2020. "The Power of Related Articles – Improving Fake News Detection on Social Media Platforms," in *Proceedings of the 53rd Hawaii International Conference on System Sciences*, Maui, USA, pp. 6063-6072.
- Gimpel, H., Heger, S., Olenberger, C., and Utz, L. 2021. "The Effectiveness of Social Norms in Fighting Fake News on Social Media," *Journal of Management Information Systems* (38:1), pp. 196-221.
- Gimpel, H., Heger, S., and Wöhl, M. 2022. "Sustainable Behavior in Motion: Designing Mobile Eco-Driving Feedback Information Systems," *Information Technology and Management* (23:4), pp. 299-314.
- Gimpel, H., Nissen, M., and Görlitz, R. 2013. "Quantifying the Quantified Self: A Study on the Motivations of Patients to Track Their Own Health," in *Proceedings of the 34th International Conference on Information Systems*, Milan, Italy, pp. 1-16.
- Gnewuch, U., Morana, S., and Maedche, A. 2017. "Towards Designing Cooperative and Social Conversational Agents for Customer Service," in *Proceedings of the 38th International Conference on Information Systems*, Seoul, South Korea.
- Goldstein, N. J., Cialdini, R. B., and Griskevicius, V. 2008. "A Room with a Viewpoint: Using Social Norms to Motivate Environmental Conservation in Hotels," *Journal of Consumer Research* (35:3), pp. 472-482.
- Gooty, J., Connelly, S., Griffith, J., and Gupta, A. 2010. "Leadership, Affect and Emotions: A State of the Science Review," *The Leadership Quarterly* (21:6), pp. 979-1004.
- Gottfried, J., and Shearer, E. 2016. "News Use Across Social Media Platforms," available at https://assets.pewresearch.org/wp-content/uploads/sites/13/2016/05/PJ_2016.05.26_social-media-and-news_FINAL-1.pdf, accessed on Feb 21 2020.
- Gould, S. J. J., Cox, A. L., and Brumby, D. P. 2016. "Diminished Control in Crowdsourcing," *ACM Transactions on Computer-Human Interaction* (23:3), pp. 1-29.
- Grady, R. H., Ditto, P. H., and Loftus, E. F. 2021. "Nevertheless, Partisanship Persisted: Fake News Warnings Help Briefly, but Bias Returns With Time," *Cognitive Research: Principles and Implications* (6:1), p. 52.

- Graen, G. B., and Uhl-Bien, M. 1995. "Relationship-Based Approach to Leadership: Development of Leader-Member Exchange (LMX) Theory of Leadership Over 25 Years: Applying a Multi-Level Multi-Domain Perspective," *The Leadership Quarterly* (6:2), pp. 219-247.
- Gratch, J., Rickel, J., André, E., Cassell, J., Petajan, E., and Badler, N. 2002. "Creating Interactive Virtual Humans: Some Assembly Required," *IEEE Intelligent Systems* (17:4), pp. 54-63.
- Gray, M. L., and Suri, S. 2019. *Ghost Work: How to Stop Silicon Valley From Building a New Global Underclass*, Boston, USA: Houghton Mifflin Harcourt.
- Greenhill, K. M., and Oppenheim, B. 2017. "Rumor Has It: The Adoption of Unverified Information in Conflict Zones," *International Studies Quarterly* (61:3), pp. 660-676.
- Gregor, S. 2006. "The Nature of Theory in Information Systems," *Management Information Systems Quarterly* (30:3), p. 611.
- Gregor, S., and Hevner, A. R. 2013. "Positioning and Presenting Design Science Research for Maximum Impact," *Management Information Systems Quarterly* (37:2), pp. 337-355.
- Groen, W. P. de, Kilhoffer, Z., Lenaerts, K., and Salez, N. 2017. "The Impact of the Platform Economy on Job Creation," *Intereconomics* (52:6), pp. 345-351.
- Guadagno, R. E., Blascovich, J., Bailenson, J. N., and McCall, C. 2007. "Virtual Humans and Persuasion: The Effects of Agency and Behavioral Realism," *Media Psychology* (10:1), pp. 1-22.
- Guthrie, S. E. 1997. "Anthropomorphism: A Definition and a Theory," in *Anthropomorphism, Anec-Dotes, and Animals*, Albany, USA: State University of New York Press.
- Guzman, D., Brito, G., Naranjo, J. E., Garcia, C. A., Saltos, L. F., and Garcia, M. V. 2018. "Virtual Assistance Environment for Deaf People Based on an Electronic Gauntlet," in *2018 IEEE Third Ecuador Technical Chapters Meeting*, Cuenca, Ecuador, pp. 1-6.
- Guzzo, R. A., Yost, P. R., Campbell, R. J., and Shea, G. P. 1993. "Potency in Groups: Articulating a Construct," *The British Journal of Social Psychology* (32:1), pp. 87-106.
- Hair, J. F., Hult, G., Ringle, C., and Sarstedt, M. 2016. *A Primer on Partial Least Squares Structural Equation Modeling (PLS-Sem)*, Newbury Park, USA: Sage Publications.
- Hair, J. F., Ringle, C., and Sarstedt, M. 2011. "PLS-SEM: Indeed a Silver Bullet," *Journal of Marketing Theory and Practice* (19:2), pp. 139-151.

- Halbesleben, J. R., Novicevic, M. M., Harvey, M. G., and Buckley, M. R. 2003. "Awareness of Temporal Complexity in Leadership of Creativity and Innovation: A Competency-Based Model," *The Leadership Quarterly* (14:4-5), pp. 433-454.
- Hammer, S., Kirchner, K., André, E., and Lugin, B. 2017. "Touch or Talk?: Comparing Social Robots and Tablet PCs for an Elderly Assistant Recommender System," in *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, Vienna, Austria, pp. 129-130.
- Häring, M., Kuchenbrandt, D., and André, E. 2014. "Would You Like to Play With Me?: How Robots' Group Membership and Task Features Influence Human-Robot Interaction," in *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, Bielefeld, Germany, pp. 9-16.
- Harms, P. D., and Credé, M. 2010. "Emotional Intelligence and Transformational and Transactional Leadership: A Meta-Analysis," *Journal of Leadership & Organizational Studies* (17:1), pp. 5-17.
- Harrison, G. W., and List, J. A. 2004. "Field Experiments," *Journal of Economic Literature* (42:4), pp. 1009-1055.
- Hasher, L., Goldstein, D., and Toppino, T. 1977. "Frequency and the Conference of Referential Validity," *Journal of Verbal Learning and Verbal Behavior* (16:1), pp. 107-112.
- Hassan, A., and Barber, S. J. 2021. "The Effects of Repetition Frequency on the Illusory Truth Effect," *Cognitive Research: Principles and Implications* (6:1), p. 38.
- Hegel, F., Krach, S., Kircher, T., Wrede, B., and Sagerer, G. 2008. "Understanding Social Robots: A User Study on Anthropomorphism," in *17th IEEE International Symposium on Robot and Human Interactive Communication*, Munich, Germany, pp. 574-579.
- Hendriks, A.-M., Habraken, J., Jansen, M. W. J., Gubbels, J. S., Vries, N. K. de, van Oers, H., Michie, S., Atkins, L., and Kremers, S. P. J. 2014. "'Are We There Yet?' - Operationalizing the Concept of Integrated Public Health Policies," *Health Policy* (114:2-3), pp. 174-182.
- Hevner, A. R. 2021. "The Duality of Science: Knowledge in Information Systems Research," *Journal of Information Technology* (36:1), pp. 72-76.
- Hevner, A. R., and Chatterjee, S. 2010. "Design Science Research in Information Systems," in *Design Research in Information Systems*, A. R. Hevner and S. Chatterjee (eds.), Boston, USA: Springer US, pp. 9-22.

- Hevner, A. R., March, S. T., Park, J., and Ram, S. 2004. "Design Science in Information Systems Research," *Management Information Systems Quarterly* (28:1), p. 75.
- Hilbert, M. 2012. "Toward a Synthesis of Cognitive Biases: How Noisy Information Processing Can Bias Human Decision Making," *Psychological Bulletin* (138:2), pp. 211-237.
- Hirschheim, R., and Klein, H. 2012. "A Glorious and Not-So-Short History of the Information Systems Field," *Journal of the Association for Information Systems* (13:4), pp. 188-235.
- Höddinghaus, M., Sondern, D., and Hertel, G. 2021. "The Automation of Leadership Functions: Would People Trust Decision Algorithms?" *Computers in Human Behavior* (116).
- Horner, C. G., Galletta, D., Crawford, J., and Shirsat, A. 2021. "Emotions: The Unexplored Fuel of Fake News on Social Media," *Journal of Management Information Systems* (38:4), pp. 1039-1066.
- Housholder, E. E., and LaMarre, H. L. 2014. "Facebook Politics: Toward a Process Model for Achieving Political Source Credibility Through Social Media," *Journal of Information Technology & Politics* (11:4), pp. 368-382.
- Howe, J. 2008. *Crowdsourcing*, London, UK: Random House.
- Howell, J. M., and Hall-Merenda, K. E. 1999. "The Ties That Bind: The Impact of Leader-Member Exchange, Transformational and Transactional Leadership, and Distance on Predicting Follower Performance," *The Journal of Applied Psychology* (84:5), pp. 680-694.
- Huang, L., Morency, L.-P., and Gratch, J. 2011. "Virtual Rapport 2.0," in *Intelligent Virtual Agents*, D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, H. H. Vilhjálmsson, S. Kopp, S. Marsella and K. R. Thórisson (eds.), Berlin, Germany: Springer Berlin Heidelberg, pp. 68-79.
- Huang, M.-H., and Rust, R. T. 2018. "Artificial Intelligence in Service," *Journal of Service Research* (21:2), pp. 155-172.
- Huber, P. J. 1967. "The Behavior of Maximum Likelihood Estimates Under Nonstandard Conditions," in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, USA, pp. 221-233.

- Hudlicka, E. 2003. "To Feel or Not to Feel: The Role of Affect in Human-Computer Interaction," *International Journal of Human-Computer Studies* (59:1-2), pp. 1-32.
- Hunt, E. 2017. "'Disputed by Multiple Fact-Checkers': Facebook Rolls Out New Alert to Combat Fake News," available at <https://www.theguardian.com/technology/2017/mar/22/facebook-fact-checking-tool-fake-news>, accessed on Jan 22 2019.
- Huotari, K., and Hamari, J. 2012. "Defining Gamification – A Service Marketing Perspective," in *Proceeding of the 16th International Academic MindTrek Conference*, Tampere, Finland, pp. 17-22.
- Hutto, C., and Gilbert, E. 2014. "VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text," *Proceedings of the International AAAI Conference on Web and Social Media* (8:1), pp. 216-225.
- Ihl, A., Strunk, K. S., and Fiedler, M. 2020. "The Mediated Effects of Social Support in Professional Online Communities on Crowdfunder Engagement in Micro-Task Crowdfunding," *Computers in Human Behavior* (113).
- Jacobson, R. P., Mortensen, C. R., and Cialdini, R. B. 2011. "Bodies Obligated and Unbound: Differentiated Response Tendencies for Injunctive and Descriptive Social Norms," *Journal of Personality and Social Psychology* (100:3), pp. 433-448.
- Jäger, G., Zilian, L. S., Hofer, C., and Füllsack, M. 2019. "Crowdfunding: Working With or Against the Crowd?" *Journal of Economic Interaction and Coordination* (14:4), pp. 761-788.
- Jalaliniya, S., and Pederson, T. 2015. "Designing Wearable Personal Assistants for Surgeons: An Egocentric Approach," *IEEE Pervasive Computing* (14:3), pp. 22-31.
- Jannoo, Z., Yap, B. W., Auchoybur, N., and Lazim, M. A. 2014. "The Effect of Nonnormality on CB-SEM and PLS-SEM Path Estimates," *International Journal of Mathematical, Computational, Physical, Electrical and Computer Engineering* (8:2), pp. 285-291.
- Jesse, M., and Jannach, D. 2021. "Digital Nudging With Recommender Systems: Survey and Future Directions," *Computers in Human Behavior Reports* (3), pp. 1-14.
- Jin, S.-A. A. 2009. "Avatars Mirroring the Actual Self Versus Projecting the Ideal Self: The Effects of Self-Priming on Interactivity and Immersion in an Exergame, Wii Fit," *Cyberpsychology & Behavior* (12:6), pp. 761-765.
- Johnson-Laird, P. N. 1983. *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness*, Boston, USA: Harvard University Press.

- Johnston, J. D., Massey, A. P., and DeVaneaux, C. 2012. "Innovation in Weight Loss Intervention Programs: An Examination of a 3D Virtual World Approach," in *Proceedings of the 45th Hawaii International Conference on System Sciences*, Maui, USA, pp. 2890-2899.
- Joo, Y. K., and Kim, K. 2017. "When You Exercise Your Avatar in a Virtual Game: The Role of Avatars' Body Shape and Behavior in Users' Health Behavior," *Interacting With Computers* (29:3), pp. 455-466.
- Joseph, D. L., Newman, D. A., and Sin, H.-P. 2011. "Leader–Member Exchange (LMX) Measurement: Evidence for Consensus, Construct Breadth, and Discriminant Validity," in *Building Methodological Bridges*, D. D. Bergh and D. J. Ketchen (eds.), Emerald Publishing Limited.
- Jung, D., Dorner, V., Glaser, F., and Morana, S. 2018. "Robo-Advisory: Digitalization and Automation of Financial Advisory," *Business & Information Systems Engineering* (60:1), pp. 81-86.
- Kahneman, D. 2011. *Thinking, Fast and Slow*, New York, USA: Macmillan.
- Kang, R., Dabbish, L., Fruchter, N., and Kiesler, S. 2015. "My Data Just Goes Everywhere: User Mental Models of the Internet and Implications for Privacy and Security," in *Eleventh Symposium On Usable Privacy and Security*, Ottawa, Canada, pp. 39-52.
- Kankanhalli, A., Xia, Q., and Zhao, X. 2021. "Understanding Personalization for Health Behavior Change Applications: A Review and Future Directions," *AIS Transactions on Human-Computer Interaction* (13:3), pp. 316-349.
- Kantosalo, A., and Toivonen, H. 2016. "Modes for Creative Human-Computer Collaboration: Alternating and Task-Divided Co-Creativity," in *Proceedings of the Seventh International Conference on Computational Creativity*, Paris, France, pp. 77-84.
- Kaplan, A. M., and Haenlein, M. 2010. "Users of the World, Unite! The Challenges and Opportunities of Social Media," *Business Horizons* (53:1), pp. 59-68.
- Kaplan, B., and Maxwell, J. A. 2005. "Qualitative Research Methods for Evaluating Computer Information Systems," in *Evaluating the Organizational Impact of Healthcare Information Systems*, J. G. Anderson and C. E. Aydin (eds.), New York, USA: Springer-Verlag, pp. 30-55.
- Karademas, E. C. 2006. "Self-Efficacy, Social Support and Well-Being," *Personality and Individual Differences* (40:6), pp. 1281-1290.

- Kassim, E. S., Jailani, S. F. A. K., Hairuddin, H., and Zamzuri, N. H. 2012. "Information System Acceptance and User Satisfaction: The Mediating Role of Trust," *Procedia-Social and Behavioral Sciences* (57), pp. 412-418.
- Kerne, A., Webb, A. M., Smith, S. M., Linder, R., Lupfer, N., Qu, Y., Moeller, J., and Damaraju, S. 2014. "Using Metrics of Curation to Evaluate Information-Based Ideation," *ACM Transactions on Computer-Human Interaction* (21:3), pp. 1-48.
- Kidd, C. D., and Breazeal, C. 2008. "Robots at Home: Understanding Long-Term Human-Robot Interaction," in *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Nice, France, pp. 3230-3235.
- Kilic, M., Heinrich, P., and Schwabe, G. 2015. "Coercing into Completeness in Financial Advisory Service Encounters," in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, Vancouver, Canada, pp. 1324-1335.
- Kim, A., and Dennis, A. R. 2019. "Says Who? The Effects of Presentation Format and Source Rating on Fake News in Social Media," *Management Information Systems Quarterly* (43:3), pp. 1025-1039.
- Kim, A., Moravec, P. L., and Dennis, A. R. 2019. "Combating Fake News on Social Media With Source Ratings: The Effects of User and Expert Reputation Ratings," *Journal of Management Information Systems* (36:3), pp. 931-968.
- Kim, B., Xiong, A., Lee, D., and Han, K. 2021. "A Systematic Review on Fake News Research Through the Lens of News Creation and Consumption: Research Efforts, Challenges, and Future Directions," *PLoS One* (16:12), e0260080.
- Kim, S. Y., Prestopnik, N., and Biocca, F. A. 2014. "Body in the Interactive Game: How Interface Embodiment Affects Physical Activity and Health Behavior Change," *Computers in Human Behavior* (36), pp. 376-384.
- Kim, Y., and Sundar, S. S. 2012a. "Anthropomorphism of Computers: Is It Mindful or Mindless?" *Computers in Human Behavior* (28:1), pp. 241-250.
- Kim, Y., and Sundar, S. S. 2012b. "Visualizing Ideal Self vs. Actual Self Through Avatars: Impact on Preventive Health Outcomes," *Computers in Human Behavior* (28:4), pp. 1356-1364.
- King, A. C., Bickmore, T. W., Campero, M. I., Pruitt, L. A., and Yin, J. L. 2013. "Employing Virtual Advisors in Preventive Care for Underserved Communities: Results From the COMPASS Study," *Journal of Health Communication* (18:12), pp. 1449-1464.

- King, A. C., Campero, M. I., Sheats, J. L., Castro Sweet, C. M., Hauser, M. E., Garcia, D., Chazaro, A., Blanco, G., Banda, J., Ahn, D. K., Fernandez, J., and Bickmore, T. W. 2020. "Effects of Counseling by Peer Human Advisors vs Computers to Increase Walking in Underserved Populations: The COMPASS Randomized Clinical Trial," *JAMA Internal Medicine* (180:11), pp. 1481-1490.
- Kiousis, S. 2002. "Interactivity: A Concept Explication," *New Media & Society* (4:3), pp. 355-383.
- Kirkpatrick, Shelley A., and Locke, E. A. 1996. "Direct and Indirect Effects of Three Core Charismatic Leadership Components on Performance and Attitudes," *Journal of Applied Psychology* (81:1), pp. 36-51.
- Kitchenham, B., and Charters, S. 2007. *Guidelines for Performing Systematic Literature Reviews in Software Engineering: Tech. Rep. EBSE 2007-001*, Keele University and Durham University Joint.
- Kittur, A. 2010. "Crowdsourcing, Collaboration and Creativity," *XRDS: Crossroads, The ACM Magazine for Students* (17:2), pp. 22-26.
- Kittur, A., Nickerson, J. V., Bernstein, M., Gerber, E., Shaw, A., Zimmerman, J., Lease, M., and Horton, J. 2013. "The Future of Crowd Work," in *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, San Antonio, USA, pp. 1301-1318.
- Klaassen, R., op den Akker, R., Lavrysen, T., and van Wissen, S. 2013a. "User Preferences for Multi-Device Context-Aware Feedback in a Digital Coaching System," *Journal on Multimodal User Interfaces* (7:3), pp. 247-267.
- Klaassen, R., op den Akker, R., and op den Akker, H. 2013b. "Feedback Presentation for Mobile Personalised Digital Physical Activity Coaching Platforms," in *Proceedings of the 6th International Conference on Pervasive Technologies Related to Assistive Environments*, Rhodes, Greece, pp. 1-8.
- Klein, H. J., and Mulvey, P. W. 1995. "Two Investigations of the Relationships among Group Goals, Goal Commitment, Cohesion, and Performance," *Organizational Behavior and Human Decision Processes* (61:1), pp. 44-53.
- Knote, R., Janson, A., Eigenbrod, L., and Söllner, M. 2018. "The What and How of Smart Personal Assistants: Principles and Application Domains for IS Research," in *Multikonferenz Wirtschaftsinformatik*, Lüneburg, Germany, pp. 1083-1094.

- Knote, R., Janson, A., Söllner, M., and Leimeister, J. M. 2019. "Classifying Smart Personal Assistants: An Empirical Cluster Analysis," in *Proceedings of the 52nd Hawaii International Conference on System Sciences*, Wailea, USA, pp. 2024-2033.
- Kormos, C., Gifford, R., and Brown, E. 2015. "The Influence of Descriptive Social Norm Information on Sustainable Transportation Behavior," *Environment and Behavior* (47:5), pp. 479-501.
- Koulouris, J., Jeffery, Z., Best, J., O'Neill, E., and Lutteroth, C. 2020. "Me vs. Super(wo)man: Effects of Customization and Identification in a VR Exergame," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, Honolulu, USA, pp. 1-17.
- Kovjanic, S., Schuh, S. C., and Jonas, K. 2013. "Transformational Leadership and Performance: An Experimental Investigation of the Mediating Effects of Basic Needs Satisfaction and Work Engagement," *Journal of Occupational and Organizational Psychology* (86:4), pp. 543-555.
- Kramer, L. L., ter Stal, S., Mulder, B. C., Vet, E. de, and van Velsen, L. 2020. "Developing Embodied Conversational Agents for Coaching People in a Healthy Lifestyle: Scoping Review," *Journal of Medical Internet Research* (22:2), e14058.
- Kreuter, M. W., Farrell, D. W., Olevitch, L. R., and Brennan, L. K. 2013. *Tailoring Health Messages: Customizing Communication With Computer Technology*, New York, USA: Routledge.
- Kroque, K. 2017. "Artificial Intelligence Is Here To Stay, But Consumer Trust Is A Must for AI in Business," available at <https://www.forbes.com/sites/kenkroque/2017/09/11/artificial-intelligence-is-here-to-stay-but-consumer-trust-is-a-must-for-ai-in-business>, accessed on Dec 14 2022.
- Kroll, R. R., Boyd, J. G., and Maslove, D. M. 2016. "Accuracy of a Wrist-Worn Wearable Device for Monitoring Heart Rates in Hospital Inpatients: A Prospective Observational Study," *Journal of Medical Internet Research* (18:9), e253.
- Krumeich, J., Bredehöft, T., Werth, D., and Loos, P. 2015. "Computer-Assisted E-Customs Transactions: Proposing a System to Support Small and Medium-Sized Enterprises in Electronically Declaring International Exports," *International Journal of Electronic Government Research* (11:1), pp. 21-38.

- Kuchenbrandt, D., Häring, M., Eichberg, J., Eyssel, F., and André, E. 2014. "Keep an Eye on the Task! How Gender Typicality of Tasks Influence Human–Robot Interactions," *International Journal of Social Robotics* (6:3), pp. 417-427.
- Kuem, J., Khansa, L., and Kim, S. S. 2020. "Prominence and Engagement: Different Mechanisms Regulating Continuance and Contribution in Online Communities," *Journal of Management Information Systems* (37:1), pp. 162-190.
- Kumar, M. 2022. "Using The Most Powerful AI Just Got Simpler," available at <https://www.forbes.com/sites/kmehta/2022/12/07/using-the-most-powerful-ai-just-got-simpler>, accessed on Feb 24 2023.
- Laato, S., Islam, A. K. M. N., Islam, M. N., and Whelan, E. 2020. "What Drives Unverified Information Sharing and Cyberchondria During the COVID-19 Pandemic?" *European Journal of Information Systems* (29:3), pp. 288-305.
- LaBrie, J. W., Hummer, J. F., Neighbors, C., and Larimer, M. E. 2010. "Whose Opinion Matters? The Relationship Between Injunctive Norms and Alcohol Consequences in College Students," *Addictive Behaviors* (35:4), pp. 343-349.
- Lang, A. 2000. "The Limited Capacity Model of Mediated Message Processing," *Journal of Communication* (50:1), pp. 46-70.
- Laranjo, L., Dunn, A. G., Tong, H. L., Kocaballi, A. B., Chen, J., Bashir, R., Surian, D., Gallego, B., Magrabi, F., Lau, A. Y. S., and Coiera, E. 2018. "Conversational Agents in Healthcare: A Systematic Review," *Journal of the American Medical Informatics Association* (25:9), pp. 1248-1258.
- Latané, B., and Darley, J. M. 1970. *The Unresponsive Bystander: Why Doesn't He Help?* New York, USA: Appleton-Century Crofts.
- Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C. R., Thorson, E. A., Watts, D. J., and Zittrain, J. L. 2018. "The Science of Fake News," *Science* (359:6380), pp. 1094-1096.
- Lee, A. 1999. "Inaugural Editor's Comments," *Management Information Systems Quarterly* (23:1), pp. v-xi.
- Lee, C. F., Lee, J. C., and Lee, A. C. 2000. *Statistics for Business and Financial Economics*, Singapore: World Scientific.

- Lee, S., and Choi, J. 2017. "Enhancing User Experience With Conversational Agent for Movie Recommendation: Effects of Self-Disclosure and Reciprocity," *International Journal of Human-Computer Studies* (103), pp. 95-105.
- Legler, R., and Reischl, T. 2003. "The Relationship of Key Factors in the Process of Collaboration: A Study of School-to-Work Coalitions," *The Journal of Applied Behavioral Science* (39:1), pp. 53-72.
- Lehto, T., Oinas-Kukkonen, H., and Drozd, F. 2012. "Factors Affecting Perceived Persuasiveness of a Behavior Change Support System," in *Proceedings of the 33rd International Conference on Information Systems*, Orlando, USA.
- Levi, L. 2017. "Real Fake News and Fake Fake News," *First Amendment Law Review* (16), pp. 232-327.
- Levy, Y., and Ellis, T. J. 2006. "A Systems Approach to Conduct an Effective Literature Review in Support of Information Systems Research," *Informing Science Journal* (9), pp. 181-212.
- Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., and Cook, J. 2012. "Misinformation and Its Correction: Continued Influence and Successful Debiasing," *Psychological Science in the Public Interest: A Journal of the American Psychological Society* (13:3), pp. 106-131.
- Lewandowsky, S., and van der Linden, S. 2021. "Countering Misinformation and Fake News Through Inoculation and Prebunking," *European Review of Social Psychology* (32:2), pp. 348-384.
- Li, B. J., and Lwin, M. O. 2016. "Player See, Player Do: Testing an Exergame Motivation Model Based on the Influence of the Self Avatar," *Computers in Human Behavior* (59), pp. 350-357.
- Li, B. J., Lwin, M. O., and Jung, Y. 2014. "Wii, Myself, and Size: The Influence of Proteus Effect and Stereotype Threat on Overweight Children's Exercise Motivation and Behavior in Exergames," *Games for Health Journal* (3:1), pp. 40-48.
- Lindell, M. K., and Whitney, D. J. 2001. "Accounting for Common Method Variance in Cross-Sectional Research Designs," *Journal of Applied Psychology* (86:1), pp. 114-121.
- Lippke, S., and Ziegelmann, J. P. 2008. "Theory-Based Health Behavior Change: Developing, Testing, and Applying Theories for Evidence-Based Interventions," *Applied Psychology* (57:4), pp. 698-716.

- Lisetti, C. 2009. "Features for Culturally Appropriate Avatars for Behavior-Change Promotion in At-Risk Populations," in *Annual Review of Cybertherapy and Telemedicine 2009: Advanced Technologies in the Behavioral, Social and Neurosciences*, B. K. Wiederhold and G. Riva (eds.), IOS Press, pp. 22-26.
- Lisetti, C., Amini, R., and Yasavur, U. 2015. "Now All Together: Overview of Virtual Health Assistants Emulating Face-to-Face Health Interview Experience," *Künstliche Intelligenz* (29:2), pp. 161-172.
- Lisetti, C., Amini, R., Yasavur, U., and Rische, N. 2013. "I Can Help You Change! An Empathic Virtual Agent Delivers Behavior Change Health Interventions," *ACM Transactions on Management Information Systems* (4:4), pp. 1-28.
- Liu, Y., Cao, B., Ma, K., and Fan, J. 2021. "Improving the Classification of Call Center Service Dialogue With Key Utterances," *Wireless Networks* (27:5), pp. 3395-3406.
- Loftus, E. F. 2005. "Planting Misinformation in the Human Mind: A 30-Year Investigation of the Malleability of Memory," *Learning & Memory* (12:4), pp. 361-366.
- Lubart, T. 2005. "How Can Computers Be Partners in the Creative Process: Classification and Commentary on the Special Issue," *International Journal of Human-Computer Studies* (63:4-5), pp. 365-369.
- Ludwig, B. 2015. *Planbasierte Mensch-Maschine-Interaktion in Multimodalen Assistenzsystemen*, Berlin, Germany: Springer Berlin Heidelberg.
- Luger, E., and Sellen, A. 2016. "'Like Having a Really Bad Pa': The Gulf Between User Expectation and Experience of Conversational Agents," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, San Jose, USA, pp. 5286-5297.
- Lyles, A. A., Amresh, A., Huberty, J., Todd, M., and Lee, R. E. 2017. "A Mobile, Avatar-Based App for Improving Body Perceptions Among Adolescents: A Pilot Test," *JMIR Serious Games* (5:1), e4.
- Maasberg, M., Ayaburi, E., Liu, C. Z., and Au, Y. A. 2018. "Exploring the Propagation of Fake Cyber News: An Experimental Approach," in *Proceedings of the 51st Hawaii International Conference on System Sciences*, Waikoloa Village, USA, pp. 3717-3726.
- MacCoun, R. J., Kier, E., and Belkin, A. 2006. "Does Social Cohesion Determine Motivation in Combat?: An Old Question with an Old Answer," *Armed Forces & Society* (32:4), pp. 646-654.

- MacDorman, K. F. 2006. "Subjective Ratings of Robot Video Clips for Human Likeness, Familiarity, and Eeriness: An Exploration of the Uncanny Valley," in *ICCS/CogSci-2006 Long Symposium: Toward Social Mechanisms of Android Science*, Vancouver, Canada.
- Maedche, A., Legner, C., Benlian, A., Berger, B., Gimpel, H., Hess, T., Hinz, O., Morana, S., and Söllner, M. 2019. "AI-Based Digital Assistants," *Business & Information Systems Engineering* (61:4), pp. 535-544.
- Maedche, A., Morana, S., Schacht, S., Werth, D., and Krumeich, J. 2016. "Advanced User Assistance Systems," *Business & Information Systems Engineering* (58:5), pp. 367-370.
- Maher, C. A., Davis, C. R., Curtis, R. G., Short, C. E., and Murphy, K. J. 2020. "A Physical Activity and Diet Program Delivered by Artificially Intelligent Virtual Health Coach: Proof-of-Concept Study," *JMIR mHealth and uHealth* (8:7), e17558.
- Maksl, A., Ashley, S., and Craft, S. 2015. "Measuring News Media Literacy," *Journal of Media Literacy Education* (6:3), pp. 29-45.
- Malhotra, N. K., Kim, S. S., and Patil, A. 2006. "Common Method Variance in IS Research: A Comparison of Alternative Approaches and a Reanalysis of Past Research," *Management Science* (52:12), pp. 1865-1883.
- Margolin, D. B., Hannak, A., and Weber, I. 2018. "Political Fact-Checking on Twitter: When Do Corrections Have an Effect?" *Political Communication* (35:2), pp. 196-219.
- Mark, B. G., Gualtieri, L., Rauch, E., Rojas, R., Buakum, D., and Matt, D. T. 2019. "Analysis of User Groups for Assistance Systems in Production 4.0," in *2019 IEEE International Conference on Industrial Engineering and Engineering Management*, Macao, China, pp. 1260-1264.
- Marks, D. F. 2020. "The COM-B System of Behaviour Change: Properties, Problems and Prospects," *Qeios*, pp. 1-6.
- Mavridis, P., Huang, O., Qiu, S., Gadiraju, U., and Bozzon, A. 2019. "Chatterbox: Conversational Interfaces for Microtask Crowdsourcing," in *Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization*, Larnaca, Cyprus, pp. 243-251.
- Mazzotta, I., Novielli, N., and Carolis, B. de. 2009. "Are ECAs More Persuasive Than Textual Messages?" in *Intelligent Virtual Agents*, Z. Ruttkay, M. Kipp, A. Nijholt and H. H. Vilhjálmsson (eds.), Berlin, Germany: Springer, pp. 527-528.

- McGinty, L., and Smyth, B. 2006. "Adaptive Selection: An Analysis of Critiquing and Preference-Based Feedback in Conversational Recommender Systems," *International Journal of Electronic Commerce* (11:2), pp. 35-57.
- McGrath, J. E. 1981. "Dilemmatics: The Study of Research Choices and Dilemmas," *American Behavioral Scientist* (25:2), pp. 179-210.
- McKnight, D. H., Choudhury, V., and Kacmar, C. 2002. "Developing and Validating Trust Measures for e-Commerce: An Integrative Typology," *Information Systems Research* (13:3), pp. 334-359.
- McLean, G., and Osei-Frimpong, K. 2019. "Hey Alexa ... Examine the Variables Influencing the Use of Artificial Intelligent In-Home Voice Assistants," *Computers in Human Behavior* (99), pp. 28-37.
- Mettler, T., Bächle, M., Daurer, S., and Judt, A. 2017. "Parental Control Reversed: Using ADR for Designing a Low-Cost Monitoring System for Elderly," in *Proceedings of the 38th International Conference on Information Systems*, Seoul, South Korea, pp. 1-19.
- Metzger, M. J., and Flanagin, A. J. 2013. "Credibility and Trust of Information in Online Environments: The Use of Cognitive Heuristics," *Journal of Pragmatics* (59), pp. 210-220.
- Meyer von Wolff, R., Hobert, S., Masuch, K., and Schuhmann, M. 2020. "Chatbots at Digital Workplaces – A Grounded-Theory Approach for Surveying Application Areas and Objectives," *Pacific Asia Journal of the Association for Information Systems* (12:2), pp. 1-39.
- Michie, S., Atkins, L., and West, R. 2014. *The Behaviour Change Wheel: A Guide to Designing Interventions*, Silverback Publishing.
- Michie, S., van Stralen, M. M., and West, R. 2011. "The Behaviour Change Wheel: A New Method for Characterising and Designing Behaviour Change Interventions," *Implementation Science* (6:42), pp. 1-12.
- Michie, S., Yardley, L., West, R., Patrick, K., and Greaves, F. 2017. "Developing and Evaluating Digital Interventions to Promote Behavior Change in Health and Health Care: Recommendations Resulting From an International Workshop," *Journal of Medical Internet Research* (19:6), e232.
- Mills, J. 2019. "Improving the 1957 Version of Dissonance Theory," in *Cognitive Dissonance: Reexamining a Pivotal Theory in Psychology*, E. Harmon-Jones (ed.), Washington, USA: American Psychological Association, pp. 27-39.

- Milmo, D., and Rankin, J. 2022. "EU Raises Prospect of Big Fine or Ban if Twitter Fails to Follow New Legislation," available at <https://www.theguardian.com/technology/2022/nov/30/eu-raises-prospect-of-big-fine-or-ban-if-twitter-fails-to-follow-new-legislation>, accessed on Dec 8 2022.
- Mirbabaie, M., Stieglitz, S., Brünker, F., Hofeditz, L., Ross, B., and Frick, N. R. J. 2021. "Understanding Collaboration with Virtual Assistants – The Role of Social Identity and the Extended Self," *Business & Information Systems Engineering* (63:1), pp. 21-37.
- Mohan, S., Venkatakrishnan, A., and Hartzler, A. L. 2020. "Designing an AI Health Coach and Studying Its Utility in Promoting Regular Aerobic Exercise," *ACM Transactions on Interactive Intelligent Systems* (10:2), pp. 1-30.
- Montenegro, J. L. Z., Da Costa, C. A., and Da Rosa Righi, R. 2019. "Survey of Conversational Agents in Health," *Expert Systems with Applications* (129), pp. 56-67.
- Moon, Y. 2000. "Intimate Exchanges: Using Computers to Elicit Self-Disclosure From Consumers," *Journal of Consumer Research* (26:4), pp. 323-339.
- Morana, S., Dehling, T., Reuter-Oppermann, M., and Sunyaev, A. 2017. "User Assistance for Health Care Information Systems," in *SIG-Health Pre-ICIS Workshop*, Seoul, South Korea, pp. 1-9.
- Morana, S., Pfeiffer, J., and Adam, M. T. P. 2020. "User Assistance for Intelligent Systems," *Business & Information Systems Engineering* (62:3), pp. 189-192.
- Moravec, P. L., Kim, A., and Dennis, A. R. 2018. "Flagging Fake News: System 1 vs. System 2," in *Proceedings of the 39th International Conference on Information Systems*, San Francisco, USA.
- Moravec, P. L., Kim, A., and Dennis, A. R. 2020. "Appealing to Sense and Sensibility: System 1 and System 2 Interventions for Fake News on Social Media," *Information Systems Research* (31:3), pp. 987-1006.
- Moravec, P. L., Kim, A., Dennis, A. R., and Minas, R. K. 2022. "Do You Really Know if It's True? How Asking Users to Rate Stories Affects Belief in Fake News on Social Media," *Information Systems Research* (33:3), pp. 765-1118.
- Moravec, P. L., Minas, R. K., and Dennis, A. R. 2019. "Fake News on Social Media: People Believe What They Want to Believe When It Makes No Sense at All," *Management Information Systems Quarterly* (43:4), pp. 1343-1360.
- Mori, M. 1970. "The Uncanny Valley," *Energy* (7:4), pp. 33-35.

- Morie, J. F., Kang, S.-H., and Chance, E. 2013. "The Association of In-world Avatar Investment with Expectations of Behavioral Change," in *Distributed, Ambient, and Pervasive Interactions*, D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, N. Streitz and C. Stephanidis (eds.), Berlin, Germany: Springer, pp. 466-473.
- Morkes, J., Kernal, H. K., and Nass, C. I. 1999. "Effects of Humor in Task-Oriented Human-Computer Interaction and Computer-Mediated Communication: A Direct Test of SRCT Theory," *Human-Computer Interaction* (14:4), pp. 395-435.
- Mozur, P. 2018. "A Genocide Incited on Facebook, With Posts From Myanmar's Military," available at <https://www.nytimes.com/2018/10/15/technology/myanmar-facebook-genocide.html>, accessed on Oct 13 2022.
- Mullen, B., and Copper, C. 1994. "The Relation Between Group Cohesiveness and Performance: An Integration," *Psychological Bulletin* (115:2), pp. 210-227.
- Müller-Polyzou, R., Reuter-Oppermann, M., Engbert, A., and Schmidt, R. 2021. "Identifying User Assistance Systems for Radiotherapy to Increase Efficiency and Help Saving Lives," *Health Systems* (10:4), pp. 318-336.
- Murali, P., Shamekhi, A., Parmar, D., and Bickmore, T. W. 2020. "Argumentation Is More Important Than Appearance for Designing Culturally Tailored Virtual Agents," in *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems*, Auckland, New Zealand, pp. 1940-1942.
- Mustafaraj, E., and Metaxas, P. T. 2017. "The Fake News Spreading Plague: Was it Preventable?" in *Proceedings of the 2017 ACM on Web Science Conference*, New York, USA, pp. 235-239.
- Napolitano, M. A., Hayes, S., Russo, G., Muresu, D., Giordano, A., and Foster, G. D. 2013. "Using Avatars to Model Weight Loss Behaviors: Participant Attitudes and Technology Development," *Journal of Diabetes Science and Technology* (7:4), pp. 1057-1065.
- Nass, C. I., Lombard, M., Henriksen, L., and Steuer, J. 1995. "Anthropocentrism and Computers," *Behaviour & Information Technology* (14:4), pp. 229-238.
- Nass, C. I., and Moon, Y. 2000. "Machines and Mindlessness: Social Responses to Computers," *Journal of Social Issues* (56:1), pp. 81-103.

- Nass, C. I., Moon, Y., and Carney, P. 1999. "Are People Polite to Computers? Responses to Computer-Based Interviewing Systems," *Journal of Applied Social Psychology* (29:5), pp. 1093-1109.
- Nass, C. I., Steuer, J., and Tauber, E. R. 1994. "Computers Are Social Actors," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Boston, USA.
- Navarro, J., Cebolla, A., Llorens, R., Borrego, A., and Baños, R. M. 2020a. "Manipulating Self-Avatar Body Dimensions in Virtual Worlds to Complement an Internet-Delivered Intervention to Increase Physical Activity in Overweight Women," *International Journal of Environmental Research and Public Health* (17:11), pp. 1-14.
- Navarro, J., Peña, J., Cebolla, A., and Baños, R. M. 2020b. "Can Avatar Appearance Influence Physical Activity? User-Avatar Similarity and Proteus Effects on Cardiac Frequency and Step Counts," *Health Communication* (37:2), pp. 222-229.
- Neighbors, C., Lostutter, T. W., Whiteside, U., Fossos, N., Walker, D. D., and Larimer, M. E. 2007. "Injunctive Norms and Problem Gambling Among College Students," *Journal of Gambling Studies* (23:3), pp. 259-273.
- Newlands, G., and Lutz, C. 2021. "Crowdwork and the Mobile Underclass: Barriers to Participation in India and the United States," *New Media & Society* (23:6), pp. 1341-1361.
- Newman, N., Fletcher, R., Kalogeropoulos, A., Levy, D. A. L., and Nielsen, R. K. 2017. "Reuters Institute Digital News Report 2017," available at https://reutersinstitute.politics.ox.ac.uk/sites/default/files/Digital%20News%20Report%202017%20web_0.pdf, accessed on Oct 27 2020.
- Ng, K. C., Tang, J., and Lee, D. 2021. "The Effect of Platform Intervention Policies on Fake News Dissemination and Survival: An Empirical Examination," *Journal of Management Information Systems* (38:4), pp. 898-930.
- Nigbur, D., Lyons, E., and Uzzell, D. 2010. "Attitudes, Norms, Identity and Environmental Behaviour: Using an Expanded Theory of Planned Behaviour to Predict Participation in a Kerbside Recycling Programme," *The British Journal of Social Psychology* (49:2), pp. 259-284.
- Noorbergen, T. J., Adam, M. T. P., Attia, J. R., Cornforth, D. J., and Minichiello, M. 2019. "Exploring the Design of mHealth Systems for Health Behavior Change using Mobile Biosensors," *Communications of the Association for Information Systems* (44), pp. 944-981.

- Noorbergen, T. J., Adam, M. T. P., Roxburgh, M., and Teubner, T. 2021. "Co-design in mHealth Systems Development: Insights From a Systematic Literature Review," *AIS Transactions on Human-Computer Interaction* (13:2), pp. 175-205.
- Nowak, K. 2001. "Defining and Differentiating Copresence, Social Presence and Presence As Transportation," in *Presence 2001 Conference*, Philadelphia, USA.
- Nowak, K. L. 2004. "The Influence of Anthropomorphism and Agency on Social Judgment in Virtual Environments," *Journal of Computer-Mediated Communication* (9:2).
- Nowak, K. L., and Biocca, F. A. 2003. "The Effect of the Agency and Anthropomorphism on Users' Sense of Telepresence, Copresence, and Social Presence in Virtual Environments," *Presence: Teleoperators and Virtual Environments* (12:5), pp. 481-494.
- Nowak, K. L., and Rauh, C. 2005. "The Influence of the Avatar on Online Perceptions of Anthropomorphism, Androgyny, Credibility, Homophily, and Attraction," *Journal of Computer-Mediated Communication* (11:1), pp. 153-178.
- Nowak, K. L., and Rauh, C. 2008. "Choose Your "Buddy Icon" Carefully: The Influence of Avatar Androgyny, Anthropomorphism and Credibility in Online Interactions," *Computers in Human Behavior* (24:4), pp. 1473-1493.
- Nyhan, B., and Reifler, J. 2010. "When Corrections Fail: The Persistence of Political Misperceptions," *Political Behavior* (32:2), pp. 303-330.
- Ojo, S. O., Bailey, D. P., Brierley, M. L., Hewson, D. J., and Chater, A. M. 2019. "Breaking Barriers: Using the Behavior Change Wheel to Develop a Tailored Intervention to Overcome Workplace Inhibitors to Breaking Up Sitting Time," *BMC Public Health* (19:1), pp. 1-17.
- Olafsson, S., O'Leary, T. K., and Bickmore, T. W. 2019. "Coerced Change-Talk With Conversational Agents Promotes Confidence in Behavior Change," in *Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare*, Trento, Italy, pp. 31-40.
- Olafsson, S., O'Leary, T. K., and Bickmore, T. W. 2020. "Motivating Health Behavior Change with Humorous Virtual Agents," in *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents*, Virtual Event, pp. 1-8.
- Oloronke, I., Oluwaseun, O., and Rhoda, I. 2017. "State of the Art: A Study of Human-Robot Interaction in Healthcare," *International Journal of Information Engineering and Electronic Business* (9:3), pp. 43-55.

- Olenberger, C., Schoch, M., and Utz, L. forthcoming. "Processing Fake News: The Influence of Intervention Order on the Perception of Fake News Over Time," *Under review*.
- op den Akker, H., op den Akker, R., Beinema, T., Banos, O., Heylen, D., Bedsted, B., Pease, A., Pelachaud, C., Traver Salcedo, V., Kyriazakos, S., and Hermens, H. 2018. "Council of Coaches - A Novel Holistic Behavior Change Coaching Approach," in *Proceedings of the 4th International Conference on Information and Communication Technologies for Ageing Well and e-Health*, Funchal, Portugal, pp. 219-226.
- Oppenlaender, J., Milland, K., Visuri, A., Ipeirotis, P., and Hosio, S. 2020. "Creativity on Paid Crowdsourcing Platforms," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, Honolulu, USA, pp. 1-14.
- Orji, R., Mandryk, R. L., and Vassileva, J. 2017. "Improving the Efficacy of Games for Change Using Personalization Models," *ACM Transactions on Computer-Human Interaction* (24:5), pp. 1-22.
- Oyibo, K., Adaji, I., Orji, R., Olabenjo, B., Azizi, M., and Vassileva, J. 2018. "Perceived Persuasive Effect of Behavior Model Design in Fitness Apps," in *Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization*, Singapore, pp. 219-228.
- Palvia, P., Daneshvar Kakhki, M., Ghoshal, T., Uppala, V., and Wang, W. 2015. "Methodological and Topic Trends in Information Systems Research: A Meta-Analysis of IS Journals," *Communications of the Association for Information Systems* (37), pp. 630-650.
- Palvia, P., Pinjani, P., and Sibley, E. H. 2007. "A Profile of Information Systems Research Published in Information & Management," *Information & Management* (44:1), pp. 1-11.
- Park, H. S., and Smith, S. W. 2007. "Distinctiveness and Influence of Subjective Norms, Personal Descriptive and Injunctive Norms, and Societal Descriptive and Injunctive Norms on Behavioral Intent: A Case of Two Behaviors Critical to Organ Donation," *Human Communication Research* (33:2), pp. 194-218.
- Pauleen, D. J. 2003. "An Inductively Derived Model of Leader-Initiated Relationship Building with Virtual Team Members," *Journal of Management Information Systems* (20:3), pp. 227-256.
- Pempek, T. A., Yermolayeva, Y. A., and Calvert, S. L. 2009. "College Students' Social Networking Experiences on Facebook," *Journal of Applied Developmental Psychology* (30:3), pp. 227-238.

- Peña, J., Khan, S., and Alexopoulos, C. 2016. "I Am What I See: How Avatar and Opponent Agent Body Size Affects Physical Activity Among Men Playing Exergames," *Journal of Computer-Mediated Communication* (21:3), pp. 195-209.
- Peña, J., and Kim, E. 2014. "Increasing Exergame Physical Activity Through Self and Opponent Avatar Appearance," *Computers in Human Behavior* (41), pp. 262-267.
- Peng, W. 2009. "Design and Evaluation of a Computer Game to Promote a Healthy Diet for Young Adults," *Health Communication* (24:2), pp. 115-127.
- Pennycook, G., Cannon, T. D., and Rand, D. G. 2018. "Prior Exposure Increases Perceived Accuracy of Fake News," *Journal of Experimental Psychology: General* (147:12), pp. 1865-1880.
- Pennycook, G., and Rand, D. G. 2019. "Lazy, Not Biased: Susceptibility to Partisan Fake News Is Better Explained by Lack of Reasoning Than by Motivated Reasoning," *Cognition* (188), pp. 39-50.
- Petty, R. E., and Cacioppo, J. T. 1986. "The Elaboration Likelihood Model of Persuasion," in *Advances in Experimental Social Psychology*, L. Berkowitz (ed.), San Diego, USA: Academic Press, pp. 123-205.
- Pfeuffer, N., Benlian, A., Gimpel, H., and Hinz, O. 2019. "Anthropomorphic Information Systems," *Business & Information Systems Engineering* (61:4), pp. 523-533.
- Picard, R. W., Vyzas, E., and Healey, J. 2001. "Toward Machine Emotional Intelligence: Analysis of Affective Physiological State," *IEEE Transactions on Pattern Analysis and Machine Intelligence* (23:10), pp. 1175-1191.
- Pinder, C., Vermeulen, J., Cowan, B. R., and Beale, R. 2018. "Digital Behaviour Change Interventions to Break and Form Habits," *ACM Transactions on Computer-Human Interaction* (25:3), pp. 1-66.
- Pirkkalainen, H., and Salo, M. 2016. "Two Decades of the Dark Side in the Information Systems Basket : Suggesting Five Areas For Future Research," in *Proceedings of the 24th European Conference on Information Systems*, Tel Aviv, Israel.
- Pitts, V. E., Wright, N. A., and Harkabus, L. C. 2012. "Communication in Virtual Teams: The Role of Emotional Intelligence," *Journal of Organizational Psychology* (28:1), pp. 2046-2054.
- Pratt, J. A., Hauser, K., Ugray, Z., and Patterson, O. 2007. "Looking at Human-Computer Interface Design: Effects of Ethnicity in Computer Agents," *Interacting With Computers* (19:4), pp. 512-523.

- Prochaska, J. O., and Velicer, W. F. 1997. "The Transtheoretical Model of Health Behavior Change," *American Journal of Health Promotion* (19:1), pp. 38-48.
- Provoost, S., Lau, H. M., Ruwaard, J., and Riper, H. 2017. "Embodied Conversational Agents in Clinical Psychology: A Scoping Review," *Journal of Medical Internet Research* (19:5), e151.
- Qian, Q., Yan, H., and Wang, Z. 2020. "Application Prospect of Artificial Intelligence Technology Basted on the Large Call Center," in *Proceedings of the 4th International Conference on Computer Science and Application Engineering*, Sanya, China, pp. 1-6.
- Qiu, S., Bozzon, A., Birk, M. V., and Gadiraju, U. 2021. "Using Worker Avatars to Improve Microtask Crowdsourcing," *Proceedings of the ACM on Human-Computer Interaction* (5), pp. 1-28.
- Qiu, S., Gadiraju, U., and Bozzon, A. 2020a. "Estimating Conversational Styles in Conversational Microtask Crowdsourcing," *Proceedings of the ACM on Human-Computer Interaction* (4), pp. 1-23.
- Qiu, S., Gadiraju, U., and Bozzon, A. 2020b. "Improving Worker Engagement Through Conversational Microtask Crowdsourcing," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, Honolulu, USA, pp. 1-12.
- Razali, N. M., and Yap, B. W. 2011. "Power Comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling Tests," *Journal of Statistical Modeling and Analytics* (2:1), pp. 21-33.
- Reis, J. C. S., Correia, A., Murai, F., Veloso, A., and Benevenuto, F. 2019. "Supervised Learning for Fake News Detection," *IEEE Intelligent Systems* (34:2), pp. 76-81.
- Reno, R. R., Cialdini, R. B., and Kallgren, C. A. 1993. "The Transsituational Influence of Social Norms," *Journal of Personality and Social Psychology* (64:1), pp. 104-112.
- Rimal, R. N., and Real, K. 2003. "Understanding the Influence of Perceived Norms on Behaviors," *Communication Theory* (13:2), pp. 184-203.
- Robinson, E., Fleming, A., and Higgs, S. 2014. "Prompting Healthier Eating: Testing the Use of Health and Social Norm Based Messages," *Health Psychology* (33:9), pp. 1057-1064.
- Romero, D., Bernus, P., Noran, O., Stahre, J., and Fast-Berglund, Å. 2016. "The Operator 4.0: Human Cyber-Physical Systems & Adaptive Automation Towards Human-Automation Symbiosis Work Systems," in *Advances in Production Management Systems. Initiatives for a Sustainable World*, I. Nääs, O. Vendrametto, J. Mendes Reis, R. F. Gonçalves, M. T.

- Silva, G. von Cieminski and D. Kiritsis (eds.), Springer International Publishing, pp. 677-686.
- Roozenbeek, J., and van der Linden, S. 2019. "The Fake News Game: Actively Inoculating Against the Risk of Misinformation," *Journal of Risk Research* (22:5), pp. 570-580.
- Ross, B., Heisel, J., Jung, A. K., and Stieglitz, S. 2018. "Fake News on Social Media: The (In)effectiveness of Warning Messages," in *Proceedings of the 39th International Conference on Information Systems*, San Francisco, USA, pp. 1-17.
- Rouast, P. V., Adam, M. T. P., and Chiong, R. 2021. "Deep Learning for Human Affect Recognition: Insights and New Developments," *IEEE Transactions on Affective Computing* (12:2), pp. 524-543.
- Rubin, V. L., Chen, Y., and Conroy, N. K. 2015. "Deception Detection for News: Three Types of Fakes," *Proceedings of the Association for Information Science and Technology* (52:1), pp. 1-4.
- Sah, Y. J., Ratan, R., Tsai, H.-Y. S., Peng, W., and Sarinopoulos, I. 2017. "Are You What Your Avatar Eats? Health-Behavior Effects of Avatar-Manifested Self-Concept," *Media Psychology* (20:4), pp. 632-657.
- Sally, D. 2000. "A General Theory of Sympathy, Mind-Reading, and Social Interaction, With an Application to the Prisoners' Dilemma," *Social Science Information* (39:4), pp. 567-634.
- Sánchez-Meca, J., Marín-Martínez, F., and Chacón-Moscoso, S. 2003. "Effect-Size Indices for Dichotomized Outcomes in Meta-Analysis," *Psychological Methods* (8:4), pp. 448-467.
- Sansone, C., and Morgan, C. 1992. "Intrinsic Motivation and Education: Competence in Context," *Motivation and Emotion* (16:3), pp. 249-270.
- Schacht, S., Morana, S., and Maedche, A. 2014. "The Project World - Gamification in Project Knowledge Management," in *Proceedings of the 22th European Conference on Information Systems*, Tel Aviv, Israel, pp. 1-10.
- Schmehl, K., and Lytvynenko, J. 2017. "7 Out of the 10 Most Viral Articles About Angela Merkel on Facebook Are False," available at <https://www.buzzfeednews.com/article/karstenschmehl/top-merkel-news>, accessed on Oct 7 2020.
- Schmeil, A., and Suggs, S. 2014. "'How Am I Doing?' – Personifying Health Through Animated Characters," in *Design, User Experience, and Usability. User Experience*

- Design for Everyday Life Applications and Services*, A. Marcus (ed.), Springer, pp. 91-102.
- Schmidt-Kraeplin, M., Thiebes, S., Schöbel, S., and Sunyaev, A. 2019. "Users' Game Design Element Preferences in Health Behavior Change Support Systems for Physical Activity: A Best-Worst-Scaling Approach," in *Proceedings of the 40th International Conference on Information Systems*, Munich, Germany, pp. 1-17.
- Schneeberger, T., Hladky, M., Thurner, A.-K., Volkert, J., Heimerl, A., Baur, T., André, E., and Gebhard, P. 2021. "Towards a Deeper Modeling of Emotions: The Deep Method and its Application on Shame," in *9th International Conference on Affective Computing and Intelligent Interaction*, Nara, Japan, pp. 1-8.
- Schneider, M., and Cooper, D. M. 2011. "Enjoyment of Exercise Moderates the Impact of a School-Based Physical Activity Intervention," *The International Journal of Behavioral Nutrition and Physical Activity* (8:1), pp. 1-8.
- Schuetzler, R. M., Grimes, G. M., and Scott Giboney, J. 2020. "The Impact of Chatbot Conversational Skill on Engagement and Perceived Humanness," *Journal of Management Information Systems* (37:3), pp. 875-900.
- Schulman, D., and Bickmore, T. W. 2009. "Persuading Users Through Counseling Dialogue With a Conversational Agent," in *Proceedings of the 4th International Conference on Persuasive Technology*, Claremont, USA, pp. 1-8.
- Schultz, W. P., Khazian, A. M., and Zaleski, A. C. 2008. "Using Normative Social Influence to Promote Conservation Among Hotel Guests," *Social Influence* (3:1), pp. 4-23.
- Schultz, W. P., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., and Griskevicius, V. 2007. "The Constructive, Destructive, and Reconstructive Power of Social Norms," *Psychological Science* (18:5), pp. 429-434.
- Schulz, A., Wirth, W., and Müller, P. 2020. "We Are the People and You Are Fake News: A Social Identity Approach to Populist Citizens' False Consensus and Hostile Media Perceptions," *Communication Research* (47:2), pp. 201-226.
- Schwarz, N., Sanna, L. J., Skurnik, I., and Yoon, C. 2007. "Metacognitive Experiences and the Intricacies of Setting People Straight: Implications for Debiasing and Public Information Campaigns," in *Advances in Experimental Social Psychology*, M. P. Zanna (ed.), Amsterdam, Netherlands: Elsevier, pp. 127-161.
- Seeber, I., Bittner, E., Briggs, R. O., Vreede, T. de, Vreede, G.-J. de, Elkins, A., Maier, R., Merz, A. B., Oeste-Reiß, S., Randrup, N., Schwabe, G., and Söllner, M. 2020. "Machines

- As Teammates: A Research Agenda on AI in Team Collaboration,” *Information & Management* (57:2).
- Seiderer, A., Ritschel, H., and André, E. 2020. “Development of a Privacy-By-Design Speech Assistant Providing Nutrient Information for German Seniors,” in *Proceedings of the 6th EAI International Conference on Smart Objects and Technologies for Social Good*, Antwerp, Belgium, pp. 114-119.
- Senécal, J., Loughead, T. M., and Bloom, G. A. 2008. “A Season-Long Team-Building Intervention: Examining the Effect of Team Goal Setting on Cohesion,” *Journal of Sport and Exercise Psychology* (30:2), pp. 186-199.
- Shafiei Gol, E., Stein, M.-K., and Avital, M. 2019. “Crowdwork Platform Governance Toward Organizational Value Creation,” *The Journal of Strategic Information Systems* (28:2), pp. 175-195.
- Shamir, B., House, R. J., and Arthur, M. B. 1993. “The Motivational Effects of Charismatic Leadership: A Self-Concept Based Theory,” *Organization Science* (4:4), pp. 577-594.
- Shao, C., Ciampaglia, G. L., Flammini, A., and Menczer, F. 2016. “Hoaxy: A Platform for Tracking Online Misinformation,” in *Proceedings of the 25th International Conference Companion on World Wide Web*, Montreal, Canada, pp. 745-750.
- Shi, Z. M., Lee, G. M., and Whinston, A. B. 2016. “Toward a Better Measure of Business Proximity: Topic Modeling for Industry Intelligence,” *Management Information Systems Quarterly* (40:4), pp. 1035-1056.
- Short, J., Williams, E., and Christie, B. 1976. *The Social Psychology of Telecommunications*, London, UK: Joh Wiley and Sons.
- Shu, K., Sliva, A., Wang, S., Tang, J., and Liu, H. 2017. “Fake News Detection on Social Media,” *ACM SIGKDD Explorations Newsletter* (19:1), pp. 22-36.
- Shukla, S., and Lyons, T. 2017. “Blocking Ads From Pages That Repeatedly Share False News,” available at [https://newsroom.fb.com/news/2017/08/blocking-ads-from-pages-that-repeatedly-share-false -news/](https://newsroom.fb.com/news/2017/08/blocking-ads-from-pages-that-repeatedly-share-false-news/), accessed on Jan 22 2019.
- Siemon, D. 2022. “Elaborating Team Roles for Artificial Intelligence-based Teammates in Human-AI Collaboration,” *Group Decision and Negotiation* (31:5), pp. 871-912.
- Silverman, B. G. 1992. “Human-Computer Collaboration,” *Human-Computer Interaction* (7:2), pp. 165-196.

- Silverman, C. 2016. "Here Are 50 of the Biggest Fake News Hits on Facebook From 2016," available at <https://www.buzzfeed.com/craigsilverman/top-fake-news-of-2016>, accessed on Jan 22 2019.
- Simon, H. A. 1979. "Rational Decision Making in Business Organizations," *The American Economic Review* (69:4), pp. 493-513.
- Sitzmann, T., and Yeo, G. 2013. "A Meta-Analytic Investigation of the Within-Person Self-Efficacy Domain: Is Self-Efficacy a Product of Past Performance or a Driver of Future Performance?" *Personnel Psychology* (66:3), pp. 531-568.
- Skjuve, M., Følstad, A., Fostervold, K. I., and Brandtzaeg, P. B. 2021. "My Chatbot Companion – A Study of Human-Chatbot Relationships," *International Journal of Human-Computer Studies* (149), pp. 1-14.
- Smith, A. 2016. "Gig Work, Online Selling and Home Sharing," available at <https://policycommons.net/artifacts/618180/gig-work-online-selling-and-home-sharing/1599085/>, accessed on Dec 4 2022.
- Smits, S., McCutchan, G., Wood, F., Edwards, A., Lewis, I., Robling, M., Paranjothy, S., Carter, B., Townson, J., and Brain, K. 2018. "Development of a Behavior Change Intervention to Encourage Timely Cancer Symptom Presentation Among People Living in Deprived Communities Using the Behavior Change Wheel," *Annals of Behavioral Medicine* (52:6), pp. 474-488.
- Song, H., Kim, J., Kwon, R. J., and Jung, Y. 2013. "Anti-smoking Educational Game Using Avatars As Visualized Possible Selves," *Computers in Human Behavior* (29:5), pp. 2029-2036.
- Sparr, J. L., and Sonnentag, S. 2008. "Fairness Perceptions of Supervisor Feedback, LMX, and Employee Well-Being at Work," *European Journal of Work and Organizational Psychology* (17:2), pp. 198-225.
- Spence, P. R., Lachlan, K. A., Spates, S. A., and Lin, X. 2013. "Intercultural Differences in Responses to Health Messages on Social Media From Spokespeople With Varying Levels of Ethnic Identity," *Computers in Human Behavior* (29:3), pp. 1255-1259.
- Spohr, D. 2017. "Fake News and Ideological Polarization," *Business Information Review* (34:3), pp. 150-160.
- Sternberg, R. J. 1999. "The Theory of Successful Intelligence," *Review of General Psychology* (3:4), pp. 292-316.

- Sterne, J. A. C., Savović, J., Page, M. J., Elbers, R. G., Blencowe, N. S., Boutron, I., Cates, C. J., Cheng, H.-Y., Corbett, M. S., Eldridge, S. M., Emberson, J. R., Hernán, M. A., Hopewell, S., Hróbjartsson, A., Junqueira, D. R., Jüni, P., Kirkham, J. J., Lasserson, T., Li, T., McAleenan, A., Reeves, B. C., Shepperd, S., Shrier, I., Stewart, L. A., Tilling, K., White, I. R., Whiting, P. F., and Higgins, J. P. T. 2019. "RoB 2: A Revised Tool for Assessing Risk of Bias in Randomised Trials," *BMJ* (366:14898), 1-8.
- Takarangi, M. K., Parker, S., and Garry, M. 2006. "Modernising the Misinformation Effect: The Development of a New Stimulus Set," *Applied Cognitive Psychology* (20:5), pp. 583-590.
- Tao, J., and Tan, T. 2005. "Affective Computing: A Review," in *Affective Computing and Intelligent Interaction*, D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, J. Tao, T. Tan and R. W. Picard (eds.), Berlin, Germany: Springer, pp. 981-995.
- ter Stal, S., Kramer, L. L., Tabak, M., op den Akker, H., and Hermens, H. 2020. "Design Features of Embodied Conversational Agents in eHealth: a Literature Review," *International Journal of Human-Computer Studies* (138), pp. 1-22.
- Terveen, L. G. 1995. "Overview of Human-Computer Collaboration," *Knowledge-Based Systems* (8:2-3), pp. 67-81.
- The Editorial Board. 2016. "Facebook and the Digital Virus Called Fake News," available at <https://www.nytimes.com/2016/11/20/opinion/sunday/facebook-and-the-digital-virus-called-fake-news.html>, accessed on Feb 21 2020.
- Thomas, J. G., Spitalnick, J. S., Hadley, W., Bond, D. S., and Wing, R. R. 2015. "Development of and Feedback on a Fully Automated Virtual Reality System for Online Training in Weight Management Skills," *Journal of Diabetes Science and Technology* (9:1), pp. 145-148.
- Thompson, D., Cantu, D., Rajendran, M., Rajendran, M., Bhargava, T., Zhang, Y., Chen, C., Liu, Y., and Deng, Z. 2016. "Development of A Teen-Focused Exergame," *Games for Health Journal* (5:5), pp. 342-356.
- Tickle-Degnen, L., and Rosenthal, R. 1990. "The Nature of Rapport and Its Nonverbal Correlates," *Psychological Inquiry* (1:4), pp. 285-293.

- Twitter International Company. 2018. "Q3 2018 Letter to Shareholders," available at https://s22.q4cdn.com/826641620/files/doc_financials/2018/q3/Q3-2018-Shareholder-Letter.pdf, accessed on Jan 3 2019.
- University of Oxford. 2016. "Resource For Understanding Political Bots," available at <https://www.oii.ox.ac.uk/blog/resource-for-understanding-political-bots/>, accessed on Feb 21 2020.
- Vainio, J., Korhonen, I., Kaipainen, K., Kenttä, O., and Järvinen, J. 2014. "Learning Healthy Habits With a Mobile Self-Intervention," in *Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare*, Oldenburg, Germany, pp. 150-157.
- van Breukelen, W., Schyns, B., and Le Blanc, P. 2006. "Leader-Member Exchange Theory and Research: Accomplishments and Future Challenges," *Leadership* (2:3), pp. 295-316.
- van der Linden, S., Leiserowitz, A., Rosenthal, S., and Maibach, E. 2017. "Inoculating the Public Against Misinformation About Climate Change," *Global Challenges* (1:2), pp. 1-7.
- van Knippenberg, B., and van Knippenberg, D. 2005. "Leader Self-Sacrifice and Leadership Effectiveness: The Moderating Role of Leader Prototypicality," *The Journal of Applied Psychology* (90:1), pp. 25-37.
- van Mulken, S., André, E., and Müller, J. 1998. "The Persona Effect: How Substantial Is It?" in *People and Computers XIII*, H. Johnson, L. Nigay and C. Roast (eds.), London: Springer London, pp. 53-66.
- van Mulken, S., André, E., and Müller, J. 1999. "An Empirical Study on the Trustworthiness of Life-Like Interface Agents," in *Proceedings of the HCI International'99 on Human-Computer Interaction: Communication, Cooperation, and Application Design*, Munich, Germany.
- van Vugt, H. C., Bailenson, J. N., Hoorn, J. F., and Konijn, E. A. 2008. "Effects of Facial Similarity on User Responses to Embodied Agents," *ACM Transactions on Computer-Human Interaction* (17:2), pp. 1-27.
- van Vugt, H. C., Konijn, E. A., Hoorn, J. F., and Veldhuis, J. 2006. "Why Fat Interface Characters Are Better e-Health Advisors," in *Intelligent Virtual Agents*, D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, J. Gratch, M. Young, R. Aylett, D. Ballin and P. Olivier (eds.), Berlin, Germany: Springer, pp. 1-13.

- van Vugt, H. C., Konijn, E. A., Hoorn, J. F., and Veldhuis, J. 2009. "When Too Heavy Is Just Fine: Creating Trustworthy E-health Advisors," *International Journal of Human-Computer Studies* (67:7), pp. 571-583.
- Vecchi, G. M., van Hasselt, V. B., and Romano, S. J. 2005. "Crisis (Hostage) Negotiation: Current Strategies and Issues in High-Risk Conflict Resolution," *Aggression and Violent Behavior* (10:5), pp. 533-551.
- Venkatesh, V., Brown, S. A., and Bala, H. 2013. "Bridging the Qualitative-Quantitative Divide: Guidelines for Conducting Mixed Methods Research in Information Systems," *Management Information Systems Quarterly* (37:1), pp. 21-54.
- Verplanken, B., and Wood, W. 2006. "Interventions to Break and Create Consumer Habits," *Journal of Public Policy & Marketing* (25:1), pp. 90-103.
- Vieira, A. D., Leite, H., and Volochchuk, A. V. L. 2022. "The Impact of Voice Assistant Home Devices on People With Disabilities: A Longitudinal Study," *Technological Forecasting and Social Change* (184).
- Voelpel, S. C., Eckhoff, R. A., and Förster, J. 2008. "David Against Goliath? Group Size and Bystander Effects in Virtual Knowledge Sharing," *Human Relations* (61:2), pp. 271-295.
- Volmer, J., Spurk, D., and Niessen, C. 2012. "Leader–Member Exchange (LMX), Job Autonomy, and Creative Work Involvement," *The Leadership Quarterly* (23:3), pp. 456-465.
- von der Pütten, A. M., Krämer, N. C., Gratch, J., and Kang, S.-H. 2010. "'It Doesn't Matter What You Are!' Explaining Social Effects of Agents and Avatars," *Computers in Human Behavior* (26:6), pp. 1641-1650.
- Vosoughi, S., Roy, D., and Aral, S. 2018. "The Spread of True and False News Online," *Science* (359:6380), pp. 1146-1151.
- Vtyurina, A., and Fourney, A. 2018. "Exploring the Role of Conversational Cues in Guided Task Support with Virtual Assistants," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, Montreal, Canada, pp. 1-7.
- Waddell, T. F., Sundar, S. S., and Auriemma, J. 2015. "Can Customizing an Avatar Motivate Exercise Intentions and Health Behaviors Among Those with Low Health Ideals?" *Cyberpsychology, Behavior and Social Networking* (18:11), pp. 687-690.
- Walter, N., and Tukachinsky, R. 2020. "A Meta-Analytic Examination of the Continued Influence of Misinformation in the Face of Correction: How Powerful Is It, Why Does It Happen, and How to Stop It?" *Communication Research* (47:2), pp. 155-177.

- Walters, M. L., Syrdal, D. S., Dautenhahn, K., te Boekhorst, R., and Koay, K. L. 2008. "Avoiding the Uncanny Valley: Robot Appearance, Personality and Consistency of Behavior in an Attention-Seeking Home Scenario for a Robot Companion," *Autonomous Robots* (24:2), pp. 159-178.
- Wandke, H. 2005. "Assistance in Human–Machine Interaction: A Conceptual Framework and a Proposal for a Taxonomy," *Theoretical Issues in Ergonomics Science* (6:2), pp. 129-155.
- Wang, J., Yang, Z., and Bhattacharjee, S. 2011. "Same Coin, Different Sides: Differential Impact of Social Learning on Two Facets of Music Piracy," *Journal of Management Information Systems* (28:3), pp. 343-384.
- Wang, W., and Benbasat, I. 2016. "Empirical Assessment of Alternative Designs for Enhancing Different Types of Trusting Beliefs in Online Recommendation Agents," *Journal of Management Information Systems* (33:3), pp. 744-775.
- Warzel, C., and Mac, R. 2018. "Twitter Just Launched a Midterm Elections Page and It's Already Full of Garbage," available at <https://www.buzzfeednews.com/article/charliwarzel/twitter-just-launched-a-midterms-page-and-its-already>, accessed on Jan 3 2019.
- Wathen, C. N., and Burkell, J. 2002. "Believe It or Not: Factors Influencing Credibility on the Web," *Journal of the American Society for Information Science and Technology* (53:2), pp. 134-144.
- Watson, A., Bickmore, T. W., Cange, A., Kulshreshtha, A., and Kvedar, J. 2012. "An Internet-Based Virtual Coach to Promote Physical Activity Adherence in Overweight Adults: Randomized Controlled Trial," *Journal of Medical Internet Research* (14:1).
- Watson, R. T., Pitt, L. F., and Kavan, C. B. 1998. "Measuring Information Systems Service Quality: Lessons from Two Longitudinal Case Studies," *Management Information Systems Quarterly* (22:1), p. 61.
- Wayne, S. J., Shore, L. M., Bommer, W. H., and Tetrick, L. E. 2002. "The Role of Fair Treatment and Rewards in Perceptions of Organizational Support and Leader-Member Exchange," *The Journal of Applied Psychology* (87:3), pp. 590-598.
- Weber, M., and Camerer, C. F. 1998. "The Disposition Effect in Securities Trading: An Experimental Analysis," *Journal of Economic Behavior & Organization* (33:2), pp. 167-184.

- Weedon, J., Nuland, W., and Stamos, A. 2017. "Information Operations and Facebook," available at <https://fbnewsroomus.files.wordpress.com/2017/04/facebook-and-information-operations-v1.pdf>, accessed on Jan 22 2019.
- Weinmann, M., Schneider, C., and vom Brocke, J. 2016. "Digital Nudging," *Business & Information Systems Engineering* (58:6), pp. 433-436.
- Wenzel, M. 2004. "The Social Side of Sanctions: Personal and Social Norms As Moderators of Deterrence," *Law and Human Behavior* (28:5), pp. 547-567.
- Wesche, J. S., and Sonderegger, A. 2019. "When Computers Take the Lead: The Automation of Leadership," *Computers in Human Behavior* (101), pp. 197-209.
- Wexelblat, A. 1998. "Don't Make That Face: A Report on Anthropomorphizing an Interface," *Intelligent Environments*, 173-170.
- White, H. 1980. "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica: Journal of the Econometric Society* (48:4), p. 817.
- Williams, C. 2007. "Research Methods," *Journal of Business & Economics Research* (5:3), pp. 65-71.
- Winkler, R., Neuweiler, M. L., Bittner, E., and Söllner, M. 2019. "Hey Alexa, Please Help Us Solve This Problem! How Interactions with Smart Personal Assistants Improve Group Performance," in *Proceedings of the 40th International Conference on Information Systems*, Munich, Germany.
- Wohl, M. J. A., and Branscombe, N. R. 2008. "Remembering Historical Victimization: Collective Guilt for Current Ingroup Transgressions," *Journal of Personality and Social Psychology* (94:6), pp. 988-1006.
- Wolff, S. B., Pescosolido, A. T., and Druskat, V. U. 2002. "Emotional Intelligence As the Basis of Leadership Emergence in Self-Managing Teams," *The Leadership Quarterly* (13:5), pp. 505-522.
- Wong, R. Y. M., Cheung, C. M. K., and Xiao, B. 2016. "Combating Online Abuse: What Drives People to Use Online Reporting Functions on Social Networking Sites," in *49th Hawaii International Conference on System Sciences*, Koloa, USA, pp. 415-424.
- Wong, S. I., Bunjak, A., Černe, M., and Fieseler, C. 2021. "Fostering Creative Performance of Platform Crowdworkers: The Digital Feedback Dilemma," *International Journal of Electronic Commerce* (25:3), pp. 263-286.

- World Health Organization. 2020. "Preventing Noncommunicable Diseases," available at www.who.int/activities/preventing-noncommunicable-diseases, accessed on Nov 29 2022.
- Wu, C.-H., and Liang, W.-B. 2011. "Emotion Recognition of Affective Speech Based on Multiple Classifiers Using Acoustic-Prosodic Information and Semantic Labels," *IEEE Transactions on Affective Computing* (2:1), pp. 10-21.
- Wu, Y. W., and Bailey, B. P. 2021. "Better Feedback from Nicer People: Narrative Empathy and Ingroup Framing Improve Feedback Exchange," *Proceedings of the ACM on Human-Computer Interaction* (4), pp. 1-20.
- Wyler, H., and Oswald, M. E. 2016. "Why Misinformation Is Reported: Evidence From a Warning and a Source-Monitoring Task," *Memory* (24:10), pp. 1419-1434.
- Yasavur, U., Lisetti, C., and Rishe, N. 2014. "Let's Talk! Speaking Virtual Counselor Offers You a Brief Intervention," *Journal on Multimodal User Interfaces* (8:4), pp. 381-398.
- Yee, N., and Bailenson, J. N. 2007. "The Proteus Effect: The Effect of Transformed Self-Representation on Behavior," *Human Communication Research* (33:3), pp. 271-290.
- Zaccaro, S. J., Blair, V., Peterson, C., and Zazanis, M. 1995. "Collective Efficacy," in *Self-Efficacy, Adaptation, and Adjustment*, C. R. Snyder and J. E. Maddux (eds.), Boston, MA: Springer US, pp. 305-328.
- Zeng, X., Abumansour, A. S., and Zubiaga, A. 2021. "Automated Fact-Checking: A Survey," *Language and Linguistics Compass* (15:10), pp. 1-21.
- Zhou, S., Zhang, Z., and Bickmore, T. W. 2017. "Adapting a Persuasive Conversational Agent for the Chinese Culture," in *Proceedings of the International Conference on Culture and Computing, Culture and Computing*, Kyoto, Japan, pp. 89-96.
- Zierau, N., Wambsganss, T., Janson, A., Schöbel, S., and Leimeister, J. M. 2020. "The Anatomy of User Experience with Conversational Agents: A Taxonomy and Propositions of Service Clues," in *Proceedings of the 41th International Conference on Information Systems*, Hyderabad, India, pp. 1-17.
- Złotowski, J., Proudfoot, D., Yogeewaran, K., and Bartneck, C. 2015. "Anthropomorphism: Opportunities and Challenges in Human-Robot Interaction," *International Journal of Social Robotics* (7:3), pp. 347-360.

Appendix A Basic User Assistance Systems to Combat Fake News

Appendix A.1 Detailed Sample Description

a) Sample Description Phase 1

Personal information					
Number of participants	538				
Number of participants in each treatment	Main treatments			Additional control treatments	
	Control	Ex-ante intervention	Ex-post intervention	Ex-ante warning-only	Ex-post warning-only
	111	100	106	109	112
Occupation	Employee	Self-employed	Unemployed	Other occupation	
	341	96	36	65	
Educational level	Bachelor's degree	High school degree or equivalent	Master's degree	Others	
	253	168	62	55	
Age	Mean		Standard deviation		
	44.33		12.94		
Gender	Female	Male	Others	Prefer not to reply	
	279	247	4	8	
Political orientation	Liberal		Conservative		
	304		234		
Social media usage	Several times a day	Once or several times a week	Less than once a week	Never	
	311	187	30	10	

Table 16: Detailed Sample Description of Phase 1

b) *Sample Description Phase 2*

Personal information					
Number of participants	493				
Number of participants in each treatment	Main treatments			Additional control treatments	
	Control	Ex-ante intervention	Ex-post intervention	Ex-ante warning-only	Ex-post warning-only
	101	93	98	99	102
Occupation	Employee	Self-employed	Unemployed	Other occupation	
	311	91	32	59	
Educational level	Bachelor's degree	High school degree or equivalent	Master's degree	Others	
	234	154	55	50	
Age	Mean		Standard deviation		
	44.63		12.89		
Gender	Female	Male	Others	Prefer not to reply	
	258	223	4	8	
Political orientation	Liberal		Conservative		
	275		218		
Social media usage	Several times a day	Once or several times a week	Less than once a week	Never	
	286	168	30	9	

Table 17: Detailed Sample Description of Phase 2

Appendix A.2 Results Regarding Structural Equality of Treatments

a) Results of Phase 1

Sample property	ANOVA		Levene-Test	
	F-value	p-Value	F-Value	p-Value
Age	1.452	0.215	1.520	0.195
Political orientation	0.846	0.497	0.098	0.983
Frequency of social media usage	0.075	0.990	0.075	0.990

Note: ⁺ $p < 0.1$, * $p < .05$, ** $p < .01$, *** $p < 0.001$

Sample property	χ^2 -test	
	χ^2	p-Value
Gender	16.759	0.151
Occupation	24.598	0.434
Education	18.729	0.537

Note: ⁺ $p < 0.1$, * $p < .05$, ** $p < .01$, *** $p < 0.001$

Table 18: Structural Equality of the Characteristics of the Participants in Phase 1

b) Results of Phase 2

Sample property	ANOVA		Levene-Test	
	F-value	p-Value	F-Value	p-Value
Age	1.053	0.379	2.005	0.093
Political orientation	1.300	0.269	0.098	0.983
Frequency of social media usage	0.027	0.999	0.027	0.999

Note: ⁺ $p < 0.1$, * $p < .05$, ** $p < .01$, *** $p < 0.001$

Sample property	χ^2 -test	
	χ^2	p-Value
Gender	18.100	0.098
Occupation	21.276	0.650
Education	16.958	0.677

Note: ⁺ $p < 0.1$, * $p < .05$, ** $p < .01$, *** $p < 0.001$

Table 19: Structural Equality of the Characteristics of the Participants in Phase 2

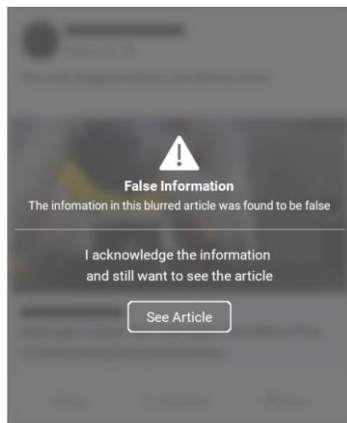
Appendix A.3 Additional Control Treatments

For additional controlling and comparing of our results with previous research, which focuses on warning labels, we used two additional control treatments:

Ex-ante warning-only: The participants are first shown a warning label and then the fake news story.

Ex-post warning-only: Participants are first shown the fake news story and then the warning label.

The task and procedure are identical to the main treatments with verified information. Only the material has been adjusted so that the news posts only contain a warning and no verified information:



(1) Ex-ante warning-only

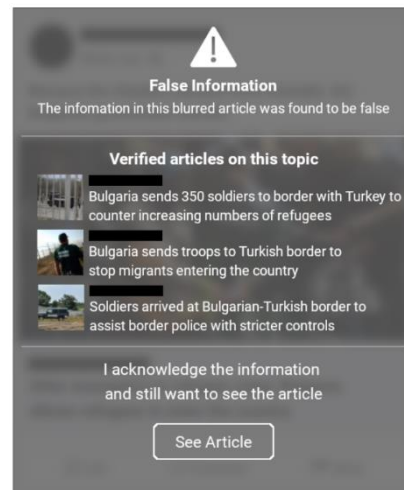
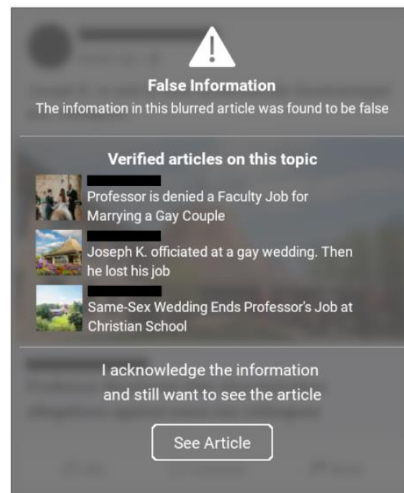
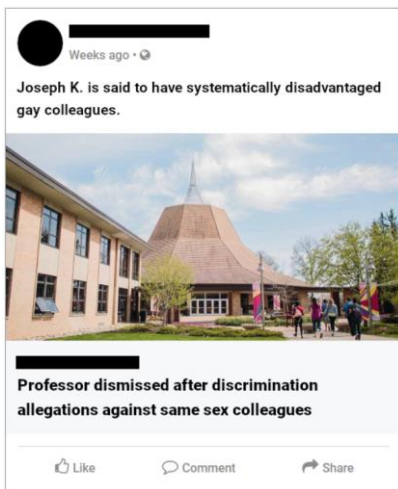


(2) Ex-post warning-only

Figure 22: Post Design in Additional Control Treatments

Appendix A.4 Stimuli

a) Fake News Posts: Fake News Posts (Left) & Corresponding Verified Information (Right)






<p>Weeks ago · 🌐</p> <p>Miranda C. and her child died in the failed medical procedure.</p>  <p>Violent riots in Croatia after failed abortion procedure</p> <p>Like Comment Share</p>	<p>False Information</p> <p>The information in this blurred article was found to be false</p> <p>Verified articles on this topic</p> <ul style="list-style-type: none"> Miranda C. denied right for abortion despite her foetus having no chance of a normal life. Abortion rights: Doctors deny Miranda C. abortion. Activists in Croatia have rallied in support of a woman who was denied an abortion. <p>I acknowledge the information and still want to see the article</p> <p>See Article</p>
<p>Weeks ago · 🌐</p> <p>The train stopped halfway and did not move.</p>  <p>Hydrogen failed: New hydrogen train Mireo Plus H failed during first presentation</p> <p>Like Comment Share</p>	<p>False Information</p> <p>The information in this blurred article was found to be false</p> <p>Verified articles on this topic</p> <ul style="list-style-type: none"> New hydrogen train Mireo Plus H successfully passes first test run Mireo Plus H hydrogen train successfully starts test runs Mireo Plus H: New hydrogen train successfully presented <p>I acknowledge the information and still want to see the article</p> <p>See Article</p>
<p>Weeks ago · 🌐</p> <p>She sold fake phone cards to her unknowing customers.</p>  <p>Woman scams charity organization out of one million egyptian pounds</p> <p>Like Comment Share</p>	<p>False Information</p> <p>The information in this blurred article was found to be false</p> <p>Verified articles on this topic</p> <ul style="list-style-type: none"> Charity goes viral: Egyptian woman raises \$1m with a phone card Egyptian woman raises \$1m for charity by selling phone card Donated phone card to be auctioned for 1 million to benefit charity <p>I acknowledge the information and still want to see the article</p> <p>See Article</p>

Figure 23: Fake News Stimuli

b) Real News Posts

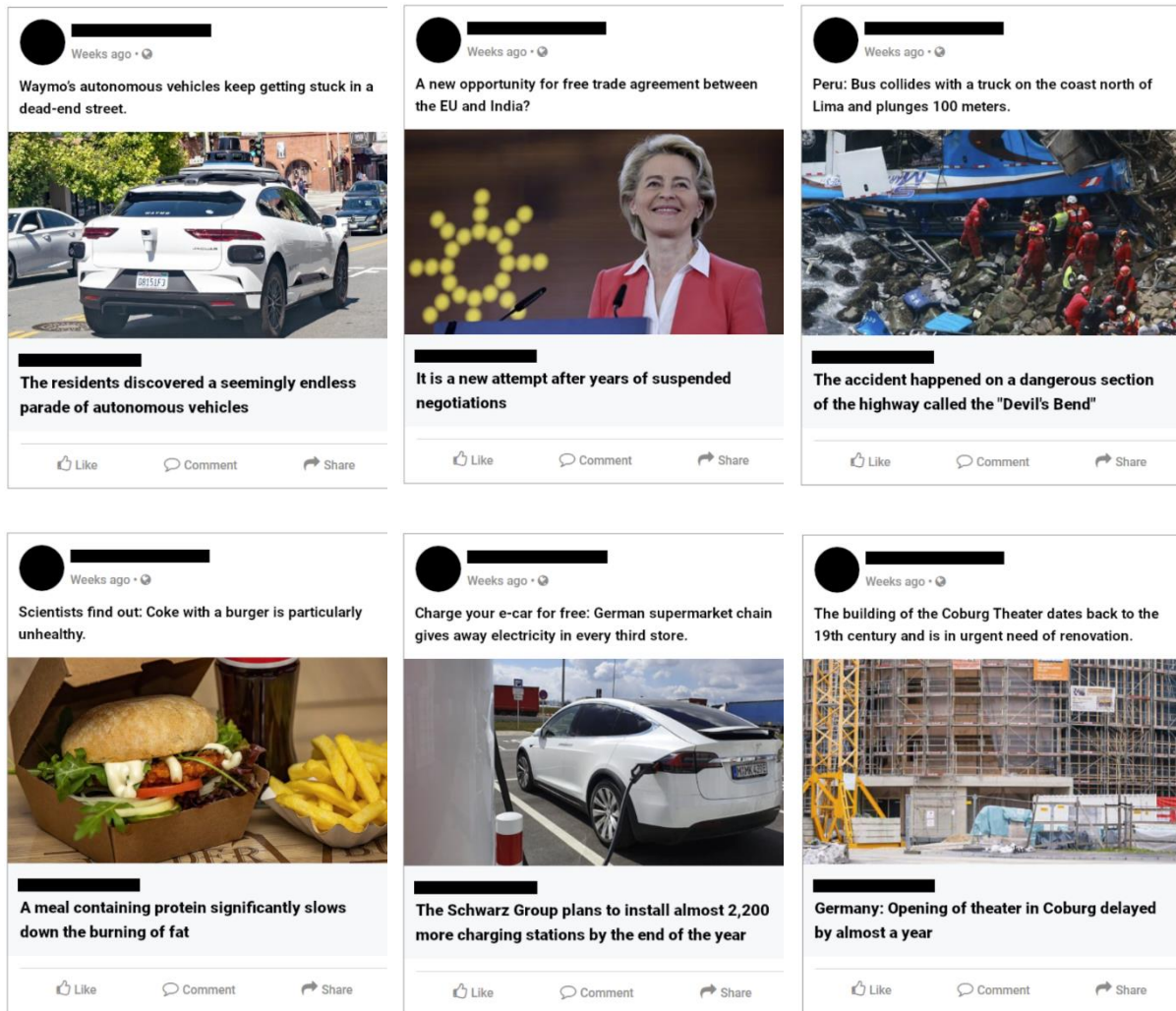


Figure 24: Real News Stimuli During Fake News Exposure

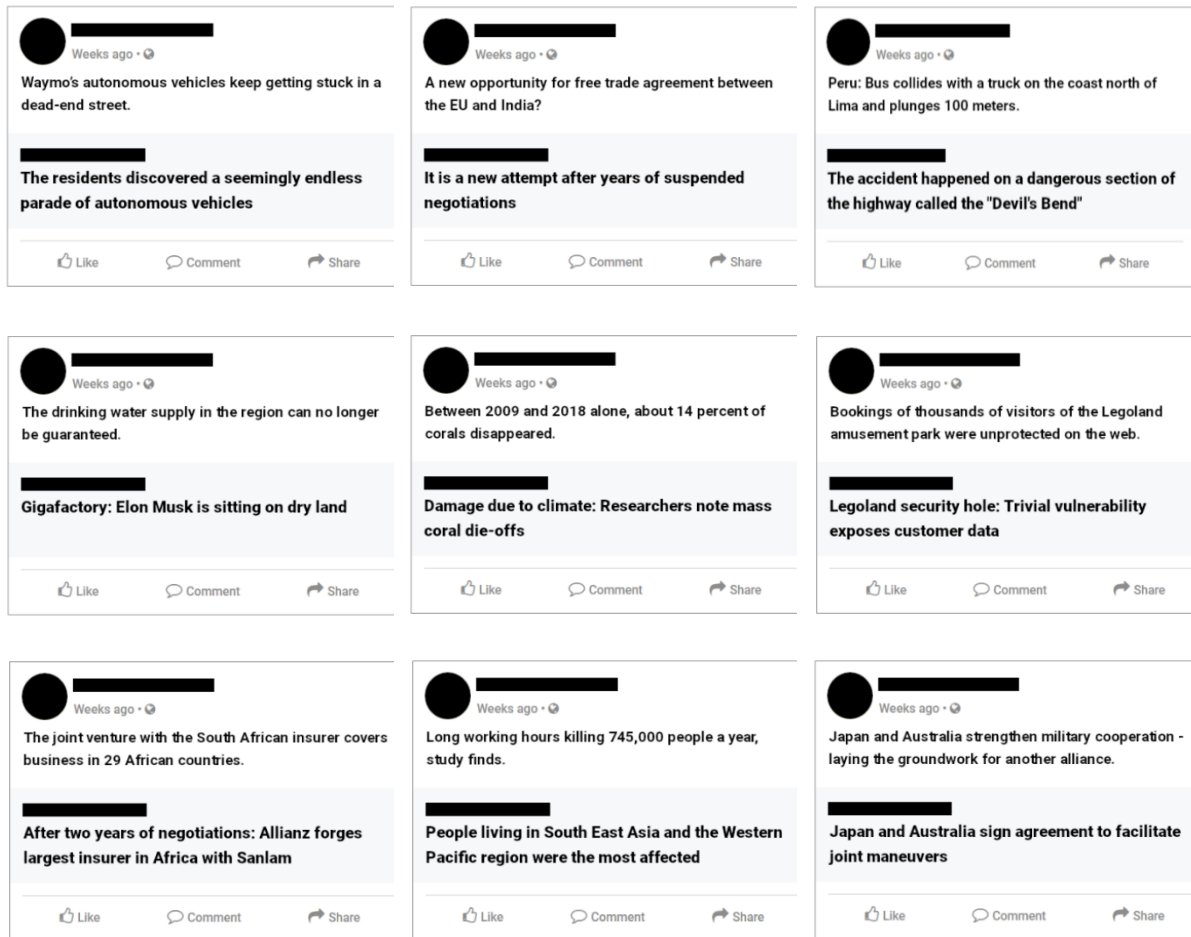


Figure 25: Real News Stimuli During Article Recognition

Appendix A.5 List of Scales

Construct	Question	Scale type	Response options	Source
Prior knowledge	How would you rate your familiarity with the following topics?: <ul style="list-style-type: none"> • Political negotiations between the European Union and India • Studies on the influence of proteins on the digestion of fat • Abortion cases in Croatia • Current development status of autonomous vehicles • Layoffs at Calvin College in Michigan • Construction projects in Coburg (Germany) • Road accidents in Peru • The installation of charging stations for electric cars in Germany • Current development status of hydrogen trains 	Single Choice	I've never heard of it, I've heard of this but I don't know what happen/ is happening, I know what happen / is happening but I don't know any specifics, I generally know about it and know some of the specifics, I am very well informed about it	Based on Kang et al. (2015)
Article familiarity	Do you remember seeing this or similar news stories before? Please exclude any previously shown information (e.g. 'verified articles') from your answer.	Single Choice	No, Yes	Based on Grady et al. (2021)
Believability	How believable do you find this article?	7-point Likert scale	Min: Not at all Max: Very much	Based on Kim and Dennis (2019)
	How truthful do you find this article?			
	How credible do you find this article?			
Confirmation Bias	Do you find the narrative described in the currently shown post important?	7-point Likert scale	Min: Not at all Max: Extremely	Based on Kim and Dennis (2019)
	What is your position on the narrative in the currently shown news post?	7-point Likert scale	Min: Extremely negative Max: Extremely positive	
Mental Model (Professor dismissed)	“I think there was an incident with a professor at Calvin College in Michigan. Did you hear about that? “ (How would you respond to this statement? Base your answer on information provided in the first session. Please choose the answer that comes closest to your personal opinion:)	Single Choice	<ul style="list-style-type: none"> • He officiated a same-sex marriage and has therefore lost his job • There were accusations of discrimination from his gay colleagues, so he was fired • He married a same-sex couple and was honored by Calvin College for doing so • Despite allegations of discrimination from his gay colleagues, he was promoted to head of the college • I have not received credible information on this topic 	Based on Takarangi et al. (2006)

Appendix A - Basic User Assistance Systems to Combat Fake News

Construct	Question	Scale type	Response options	Source
Mental Model (Border control)	<p>"I heard that the situation on the border between Turkey and Bulgaria has changed. Do you know what happened there?" (How would you respond to this statement? Base your answer on information provided in the first session. Please choose the answer that comes closest to your personal opinion:)</p>	Single Choice	<ul style="list-style-type: none"> • The number of refugees on the border with Turkey has increased, so the Bulgarian government has sent troops to stop the refugees • The situation in the refugee camps has gotten worse, so the Bulgarian government has opened the borders • As more and more refugees came, the government built a fence to stop the refugees • After a protest of the refugees the government has opened the borders • I have not received credible information on this topic 	Based on Takarangi et al. (2006)
Mental Model (Demonstrations in Croatia)	<p>"I have noticed that people in Croatia are currently demonstrating on the issue of abortions. Do you know what that's about?" (How would you respond to this statement? Base your answer on information provided in the first session. Please choose the answer that comes closest to your personal opinion:)</p>	Single Choice	<ul style="list-style-type: none"> • A woman was denied a legal abortion. People demonstrate against the unfair treatment and to support the woman • A woman died during an abortion. Therefore, people demonstrate against abortions • After a woman was denied an abortion, people are demonstrating against the abortion to save the life of the child • After a woman died during an abortion, people are demonstrating for stricter investigations of abortion clinics to make the procedures safer • I have not received credible information on this topic 	Based on Takarangi et al. (2006)
Mental Model (City Council election)	<p>"The Des Moines City Council election in Iowa took place recently. In this context, something sensational happened. Did you hear anything about it?" (How would you respond to this statement? Base your answer on information provided in the first session. Please choose the answer that comes closest to your personal opinion:)</p>	Single Choice	<ul style="list-style-type: none"> • Black Lives Matter activist Indira S. ran on a platform that included defunding the police won a seat on the Des Moines City Council • Black Lives Matter activist Indira S. was caught with drugs and therefore withdrew her candidacy • Black Lives Matter activist Indira S. campaigned on defunding the police and suffered a bitter defeat in the elections • Black Lives Matter activist Indira S. was caught with drugs, so she was expelled from her political party • I have not received credible information on this topic 	Based on Takarangi et al. (2006)

Appendix A - Basic User Assistance Systems to Combat Fake News

Construct	Question	Scale type	Response options	Source
Mental Model (Hydrogen train)	"I heard that a hydrogen train was recently introduced. Do you know how the presentation went?" (How would you respond to this statement? Base your answer on information provided in the first session. Please choose the answer that comes closest to your personal opinion:)	Single Choice	<ul style="list-style-type: none"> The Mireo Plus H was presented and has successfully completed its first test runs The Mireo Plus H was presented, but there was a technical problem and the train stopped halfway The Mireo Plus H was unveiled. However, it was only a presentation of the design without a test drive The Mireo Plus H was introduced, however, the train had a problem and could not start I have not received credible information on this topic 	Based on Takarangi et al. (2006)
Mental Model (Charity campaign)	"In Egypt, it seems that a charity campaign drew a lot of attention. Do you know what happened there?" (How would you respond to this statement? Base your answer on information provided in the first session. Please choose the answer that comes closest to your personal opinion:)	Single Choice	<ul style="list-style-type: none"> A woman sold a phone card for charity. The campaign went viral and 1 million Egyptian pounds were donated Fake phone cards were sold to a charity organization, resulting in the theft of one million Egyptian pounds A woman donated phone cards to a charity. The phone card was used for gambling, through which the charity won one million Egyptian pounds A woman sold fake phone cards for a good cause. However, the fraud was discovered and one million Egyptian pounds were recovered I have not received credible information on this topic 	Based on Takarangi et al. (2006)
Mental Model (autonomous cars)	"The company Waymo has been testing its autonomous cars for some time. Do you know whether everything works?" (How would you respond to this statement? Base your answer on information provided in the first session. Please choose the answer that comes closest to your personal opinion:)	Single Choice	<ul style="list-style-type: none"> The autonomous cars repeatedly get stuck in a dead-end street The autonomous cars drive flawlessly and without any known problems There was an incident in which a pedestrian was hit The cars have problems at night and can be used only during the day I have not received credible information on this topic 	Based on Takarangi et al. (2006)
Mental Model (Negotiations EU)	"I have heard about negotiations between the European Union and India. Do you know what they talked about?" (How would you respond to this statement? Base your answer on information provided in the first session. Please choose the answer that comes closest to your personal opinion:)	Single Choice	<ul style="list-style-type: none"> The European Union and India discussed the improvement of cultural exchange There was a new attempt to sign a free trade agreement The European Union and India discussed arms exports from the EU The European Union has tried to act as a mediator in the conflict between India and Pakistan I have not received credible information on this topic 	Based on Takarangi et al. (2006)

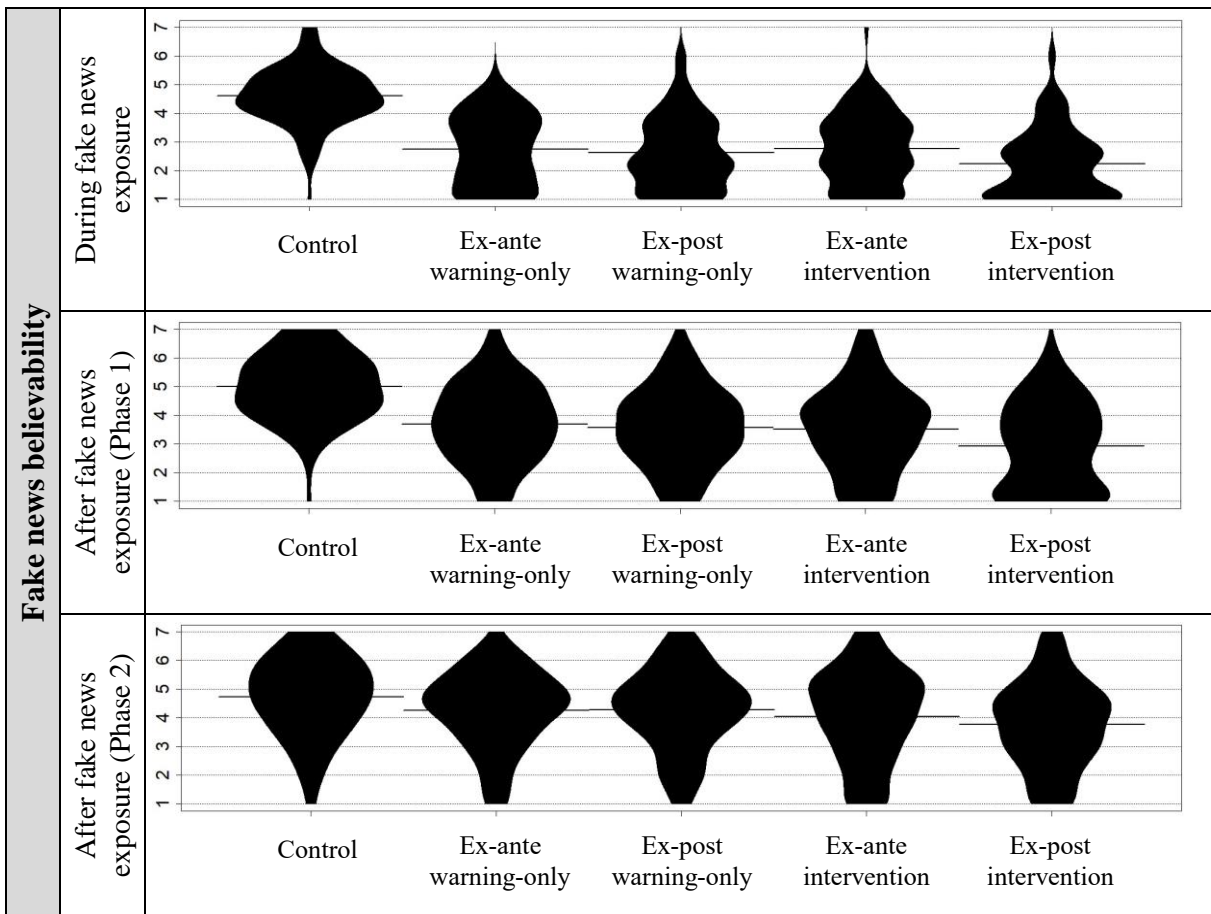
Appendix A - Basic User Assistance Systems to Combat Fake News

Construct	Question	Scale type	Response options	Source
Mental Model (Accident Peru)	"There was a terrible accident in Peru. Do you know anything about it?" (How would you respond to this statement? Base your answer on information provided in the first session. Please choose the answer that comes closest to your personal opinion:)	Single Choice	<ul style="list-style-type: none"> A tree on the tracks caused a train to derail A bus collided with a truck and fell down a cliff There was a plane crash near Lima There was a mass collision of cars near Lima I have not received credible information on this topic 	Based on Takarangi et al. (2006)
Article recognition	Do you remember seeing his exact news article (with an image) during the first session (approx. two weeks ago)?	5-point Likert scale	Definitely no, I don't think so, Unsure, I think so, Definitely yes	Self-developed
	Do you remember seeing a warning message (during the first session) stating the information in this article was found to be false?			
	Do you remember seeing verified articles (during the first session) that provided more information about the topic in the article?			
Trust in medical therapies (Control variable)	I do not trust any classical and conventional medical therapies	7-point Likert scale	Min: Strongly disagree Max: Strongly agree	Based on Gimpel et al. (2013)
	I want to be independent from classical and conventional medical therapies			
Description: Credibility assessment (Control variable)	Please describe your approach used to assess the credibility of the articles in this experiment: (Please note that only meaningful answers will be accepted.)	Text	-	Self-developed
Gender	What is your gender?	Single Choice	Female, Male, Others, Prefer not to reply	Based on the standard demographic questions of the Sosci Survey tool (soscisurvey.de)
Age	Numerical input	Integer value equal or greater than 0	Numerical input	
Education	What is your highest educational achievement? (Please select the highest level of qualification you have obtained.)	Single Choice	None (e.g. still in school), Less than high school diploma, High school degree or equivalent, Bachelor's degree, Master's degree, Doctorate (e.g. Ph.D.) or higher, Other (please specify):	
Occupation	What do you do professionally?	Single Choice	Pupil/in school, Training/apprenticeship, University student, Employee, Civil servant, Self-employed, Unemployed/seeking employment, Other:	
Political orientation	How would you assess your political orientation? Politically, I would say I am...	8-point Likert scale	Min: Very liberal Max: Very conservative	
Social media usage	How often do you use social media in your everyday life?	6-point Likert scale	Never; Less than once a week; Once a week; Several times a week; Once a day; Several times a day	Self-developed (see chapter 3.3)

Table 20: Description of All Scales Used

Appendix A.6 Descriptive Overview With All Treatments

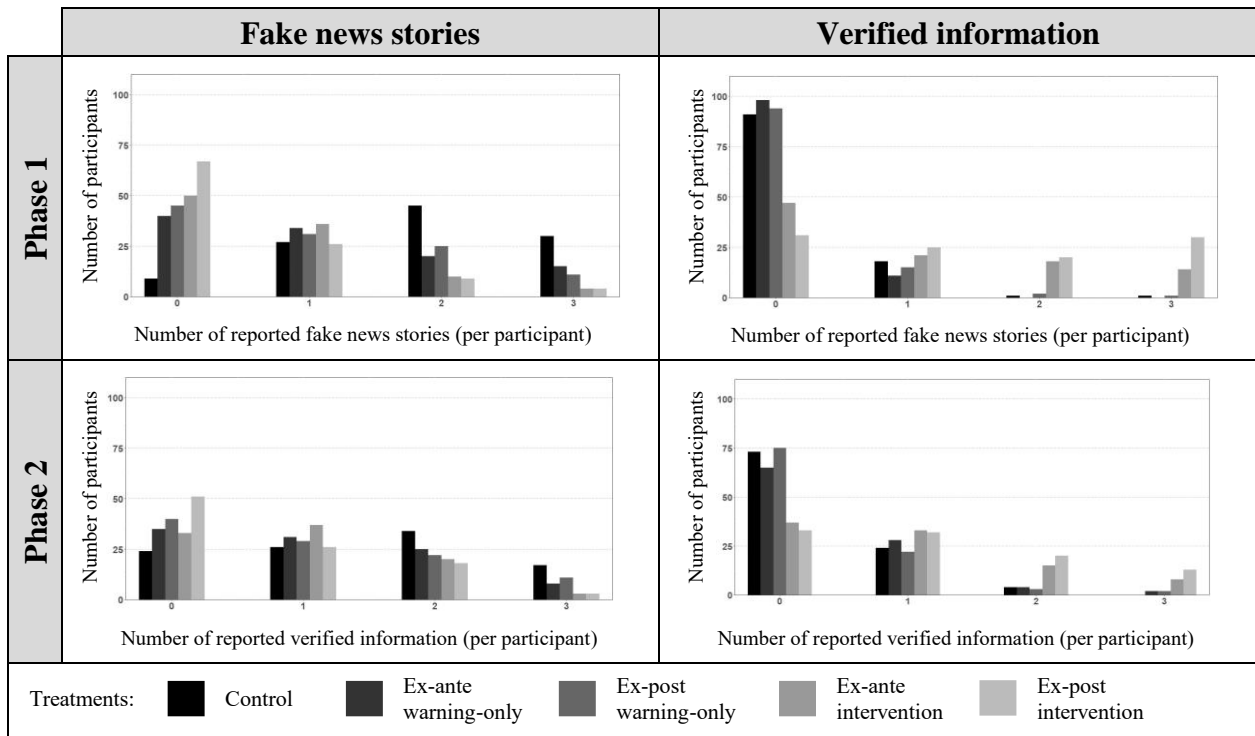
a) Comparison of the Mean Believability of Fake News



*Note: Mean value of the distribution is shown as a horizontal line.
1 - low believability, 7 - high believability*

Figure 26: Comparison of the Mean Believability of Fake News

b) Comparison of the Mental Model Measurement per Treatment



Note: Numbers vary from zero (the participant adapted no fake news story/verified information into the mental model) to three (the participant adapted all fake news stories/verified information into the mental model).

Figure 27: Mental Model Measurement per Treatment

Appendix A.7 Descriptive Overview of the Data

a) Mean Believability Measurement for Fake News Posts

Metric	Characteristic	Main treatments			Additional control treatments	
		Control treatment	Ex-ante intervention	Ex-post intervention	Ex-ante warning-only	Ex-post warning-only
Believability during fake news exposure (Phase 1a)	Mean	4.621	2.778	2.235	2.761	2.628
	Median	4.556	2.667	2.167	2.778	2.333
	Standard Deviation	0.999	1.247	1.198	1.234	1.258
Believability over time (Phase 1b)	Mean	5.015	3.521	2.932	3.700	3.570
	Median	5.000	3.778	3.000	3.667	3.611
	Standard Deviation	1.152	1.486	1.529	1.394	1.372
Believability over time (Phase 2)	Mean	4.735	4.050	3.766	4.270	4.277
	Median	4.889	4.111	3.944	4.556	4.500
	Standard Deviation	1.391	1.604	1.521	1.338	1.417

Table 21: Comparison of Mean Believability Measurement for Fake News Posts

b) Number of Fake News Adopted Into Participant's Mental Models

Metric	Characteristic	Main treatments			Additional control treatments	
		Control treatment	Ex-ante intervention	Ex-post intervention	Ex-ante warning-only	Ex-post warning-only
Number of fake news adopted into mental models (Phase 1b)	Mean	1.865	0.680	0.528	1.092	1.018
	Median	2.000	0.500	0.000	1.000	1.000
	Standard Deviation	0.909	0.815	0.807	1.050	1.013
Number of fake news adopted into mental models (Phase 2)	Mean	1.436	0.925	0.725	1.061	1.039
	Median	2.000	1.000	0.000	1.000	1.000
	Standard Deviation	1.034	0.837	0.871	0.967	1.024

Note: Results are based on mean, median, standard deviation per participant.

Table 22: Comparison of Number of Mental Models Based on Fake News Posts

c) *Number of Mental Models Based on Verified Information*

Metric	Characteristic	Main treatments			Additional control treatments	
		Control treatment	Ex-ante intervention	Ex-post intervention	Ex-ante warning-only	Ex-post warning-only
Number of verified information adopted into mental models (Phase 1b)	Mean	0.207	0.990	1.462	0.101	0.196
	Median	0.000	1.000	1.000	0.000	0.000
	Standard Deviation	0.488	1.105	1.189	0.303	0.499
Number of verified information adopted into mental models (Phase 2)	Mean	0.317	0.935	1.133	0.434	0.333
	Median	0.000	1.000	1.000	0.000	0.000
	Standard Deviation	0.546	0.953	1.032	0.671	0.634

Note: Results are based on mean, median, standard deviation per participant.

Table 23: Comparison of Number of Mental Models Based on Verified Information

Appendix A.8 Detailed Report of Results

a) Comparison of Believability Measurement for Fake News Posts Between the Treatments

Metric	Compared treatments	Wilcoxon's W	p-value	Related Hypothesis
Believability during fake news exposure (Phase 1a)	Control vs. ex-ante intervention	9765.6	< 0.001 ***	-
	Control vs. ex-post intervention	10868	< 0.001 ***	-
	Ex-ante intervention vs. ex-post intervention	6684	< 0.001 ***	H 3.2.1
	Ex-ante intervention vs. ex-ante warning-only	5459.5	0.492	-
	Ex-post intervention vs. ex-post warning-only	7043.5	0.008 **	-
Believability over time (Phase 1b)	Control vs. ex-ante intervention	8704	< 0.001 ***	-
	Control vs. ex-post intervention	9995	< 0.001 ***	-
	Ex-ante intervention vs. ex-post intervention	6398	0.005 **	H 3.2.2
	Ex-ante intervention vs. ex-ante warning-only	5827.5	0.194	-
	Ex-post intervention vs. ex-post warning-only	7324	0.001 **	-
Believability over time (Phase 2)	Control vs. ex-ante intervention	5825.5	0.002 **	-
	Control vs. ex-post intervention	6759.5	< 0.001 ***	-
	Ex-ante intervention vs. ex-post intervention	5120	0.070 ⁺	H 3.2.2
	Ex-ante intervention vs. ex-ante warning-only	4916	0.209	-
	Ex-post intervention vs. ex-post warning-only	6032	0.006 **	-

Note: Results based on one-sided Wilcoxon-Mann-Whitney pairwise comparison.

⁺ $p < 0.1$, * $p < .05$, ** $p < .01$, *** $p < 0.001$

Table 24: Results of Believability Comparison

b) Comparison of Number of Fake News Adopted Into Participant's Metal Model

Metric	Compared treatments	Wilcoxon's W	p-value	Related Hypothesis
Number of fake news adopted into mental models (Phase 1b)	Control vs. ex-ante intervention	9096	< 0.001 ***	-
	Control vs. ex-post intervention	9096	< 0.001 ***	-
	Ex-ante intervention vs. ex-post intervention	5946	0.045 *	H 3.2.3
	Ex-ante intervention vs. ex-ante warning-only	6602	0.002 **	-
	Ex-post intervention vs. ex-post warning-only	7569	< 0.001 ***	-
Number of fake news adopted into mental models (Phase 2)	Control vs. ex-ante intervention	6010.5	< 0.001 ***	-
	Control vs. ex-post intervention	6840.5	< 0.001 ***	-
	Ex-ante intervention vs. ex-post intervention	5219	0.032 *	H 3.2.3
	Ex-ante intervention vs. ex-ante warning-only	4906	0.204	-
	Ex-post intervention vs. ex-post warning-only	5829.5	0.015 *	-

Note: Results based on one-sided Wilcoxon-Mann-Whitney pairwise comparison.

+ $p < 0.1$, * $p < .05$, ** $p < .01$, *** $p < 0.001$

Table 25: Results of Fake News Adoption

c) Comparison of Number of Verified Information Adopted Into Mental Model

Metric	Compared treatments	Wilcoxon's W	p-value	Related Hypothesis
Number of verified information adopted into mental models (Phase 1b)	Control vs. ex-ante intervention	3343.5	< 0.001 ***	-
	Control vs. ex-post intervention	2350.5	< 0.001 ***	-
	Ex-ante intervention vs. ex-post intervention	4104	0.002 **	H 3.2.4
Number of verified information adopted into mental models (Phase 2)	Control vs. ex-ante intervention	2944.5	< 0.001 ***	-
	Control vs. ex-post intervention	2680.5	< 0.001 ***	-
	Ex-ante intervention vs. ex-post intervention	4084.5	0.097 ⁺	H 3.2.4

Note: Results based on one-sided Wilcoxon-Mann-Whitney pairwise comparison. The control treatment contained not verified information.

⁺ $p < 0.1$, * $p < .05$, ** $p < .01$, *** $p < 0.001$

Table 26: Results of Verified Information Adoption

d) Comparison of the Memory of the Fake News Articles

Metric	Compared treatments	Wilcoxon's W	p-value
Remember fake news article (Phase 1b)	Control vs. ex-ante intervention	5254	0.241
	Control vs. ex-post intervention	4808	0.006 **
	Ex-ante intervention vs. ex-post intervention	4593.5	0.036 *
	Ex-ante intervention vs. ex-ante warning-only	4876.5	0.085 +
	Ex-post intervention vs. ex-post warning-only	5530.5	0.166
Remember fake news article (Phase 2)	Control vs. ex-ante intervention	5452.5	0.975
	Control vs. ex-post intervention	5146.5	0.690
	Ex-ante intervention vs. ex-post intervention	4014	0.075 +
	Ex-ante intervention vs. ex-ante warning-only	4946.5	0.817
	Ex-post intervention vs. ex-post warning-only	4598	0.162

Note: Results based on one-sided Wilcoxon-Mann-Whitney pairwise comparison.

+ $p < 0.1$, * $p < .05$, ** $p < .01$, *** $p < 0.001$

Table 27: Results of Memory of the Fake News Articles

e) Comparison of the Memory of the Warning Label

Metric	Compared treatments	Wilcoxon's W	p-value
Remember warning label (Phase 1b)	Ex-ante intervention vs. ex-post intervention	4650	0.060 ⁺
	Ex-ante intervention vs. ex-ante warning-only	3858.5	< 0.001 ***
	Ex-post intervention vs. ex-post warning-only	4435.5	< 0.001 ***
Remember warning label (Phase 2)	Ex-ante intervention vs. ex-post intervention	3830	0.028 *
	Ex-ante intervention vs. ex-ante warning-only	3927	0.038 *
	Ex-post intervention vs. ex-post warning-only	3336.5	< 0.001 ***

Note: Results based on one-sided Wilcoxon-Mann-Whitney pairwise comparison. Only the treatments with warning labels are considered.
⁺ $p < 0.1$, * $p < .05$, ** $p < .01$, *** $p < 0.001$

Table 28: Results of Memory of the Warning Label

f) Comparison of the Memory of the Verified Information

Metric	Compared treatments	Wilcoxon's W	p-value
Remember verified information (Phase 1b)	Ex-ante intervention vs. ex-post intervention	4589.5	0.045 *
Remember verified information (Phase 2)	Ex-ante intervention vs. ex-post intervention	3702.5	0.012 *

Note: Results based on one-sided Wilcoxon-Mann-Whitney pairwise comparison. Only the treatments with verified information are considered.
⁺ $p < 0.1$, * $p < .05$, ** $p < .01$, *** $p < 0.001$

Table 29: Results of Memory of the Verified Information

Appendix A.9 Effects of Warning Labels Extended by Verified Information

To isolate the effect of the additional verified information, we compare the ex-ante intervention treatment with the ex-ante warning-only treatment (which contains only a warning label) and the ex-post intervention treatment with the ex-post warning-only treatment (which contains only a warning label). The results show no significant differences between the treatments regarding believability during fake news exposure. Even after fake news exposure we observe no difference between ex-ante intervention treatment and ex-ante warning-only treatment.

However, during and shortly after fake news exposure (in Phase 1) the ex-post intervention results in a significantly lower believability (p -value = 0.008 and p -value = 0.001) than the corresponding warning-only treatment. We also observe this effect in Phase 2 (p -value = 0.006).

Further analysis of the data also shows that additional verified information reduces the number of fake news adopted into the mental model. In the ex-post intervention treatment, the participants reported significantly less fake news as part of their mental model (p -value < 0.001) compared to the ex-post warning treatment. We also observe this effect after two weeks in Phase 2 (p -value = 0.015). In the ex-ante intervention treatment, we also observe that participants adopt less fake news into their mental model (p -value = 0.002) than in the ex-ante warning-only treatment. However, we can no longer observe this effect after two weeks (p -value = 0.204).

We also examine the impact on memory of the warning label itself during the intervention. We observe in Phase 1b that participants in the ex-post intervention treatment are better able to recall the presence of a warning (p -value < 0.001) than in the ex-post warning-only treatment. We continued to observe this phenomenon two weeks later in Phase 2 (p -value < 0.001). Also, in the ex-ante intervention treatment, we observe a better recall of the warning label in the first phase (p -value < 0.001) than in the ex-ante warning-only treatment. The effect is still observable after two weeks in Phase 2 (p -value = 0.038). Overall, the inclusion of additional verified information shows that participants better remember the warning labels.

Our results show that warnings as an intervention against fake news can be extended by using additional verified information. A direct comparison of the interventions with warning labels only and the interventions with additional verified information shows no difference during fake news exposure. This is plausible because the warning during the fake news exposure makes it equally obvious to the participants that the information is false. However, we see different effects over time. For example, we observe that participants adopted less fake news in their

mental model. This is consistent with our expectations. Using only a warning creates a gap in the mental model as participants have no understanding of what was actually happening. In addition, the memory of the warning fades over time (Grady et al., 2021), whereby fake news increasingly becomes part of the mental model. This gap is filled by the verified information, which makes fake news less likely to be adopted into the mental model. Furthermore, the additional verified information leads to a better memory of the warning itself.

Appendix A.10 Detailed Sample Description

a) Sample Description Study 1 Regarding the Effect of Social Norms

Personal information					
Number of participants	320				
Number of participants in each treatment	Control	Injunctive SN	Descriptive SN	Combined	
	78	83	77	82	
Occupation	Students		Employed		
	294		26		
Educational level	General qualification for university entrance	Bachelor's degree	Master's degree	Others	
	216	66	17	21	
Age	Mean		Standard deviation		
	23.2		4.16		
Gender	Male		Female		
	96		224		
Usage of social media					
Social media usage	Several times a day		Once or several times a week		
	253		67		
Trust in news available on social media	No trust and little trust	Rather little trust	Rather a lot of trust	A lot of trust and a great deal of trust	
	59	135	101	25	
Content generation compared to their social environment	No content at all	Very little content and little content	Rather little content	Rather a lot of content	A lot of content and a great deal of content
	26	201	58	28	7
Press the like button per day	Mean		Standard deviation		
	4.70		7.15		
Post comment per day	Mean		Standard deviation		
	0.91		1.96		
Share of posts per day	Mean		Standard deviation		
	0.26		1.38		
Report of posts per day	Mean		Standard deviation		
	0.12		0.51		

Table 30: Detailed Sample Description of Study 1

b) *Sample Description Study 2 Regarding the Effect of Social Norms*

Personal information					
Number of participants	157				
Number of participants in each treatment	Control	Positive Descriptive SN		Negative Descriptive SN	
	51	55		51	
Occupation	Students			Employed	
	116			33	
Educational level	General qualification for university entrance	University degree		Others	
	60	82		15	
Age	Mean			Standard deviation	
	25.6			5.21	
Gender	Male			Female	
	63			94	
Usage of social media					
Social media usage	Several times a day		Once or several times a week		
	124		33		
Trust in news available on social media	No trust and little trust	Rather little trust		Rather a lot of trust	A lot of trust and a great deal of trust
	43	59		42	13
Content generation compared to their social environment	No content at all	Very little content and little content	Rather little content	Rather a lot of content	A lot of content and a great deal of content
	16	80	41	18	2
Liked posts in comparison to the social environment	None at all	Very few and few	Rather few	Rather many	Many and a great many
	11	54	58	26	8
Commented posts in comparison to the social environment	None at all	Very few and few	Rather few	Rather many	Many and a great many
	43	72	28	12	2
Shared posts in comparison to the social environment	None at all	Very few and few	Rather few	Rather many	Many and a great many
	69	57	20	7	4
Reported posts in comparison to the social environment	None at all	Very few and few	Rather few	Rather many	Many and a great many
	105	31	9	11	1

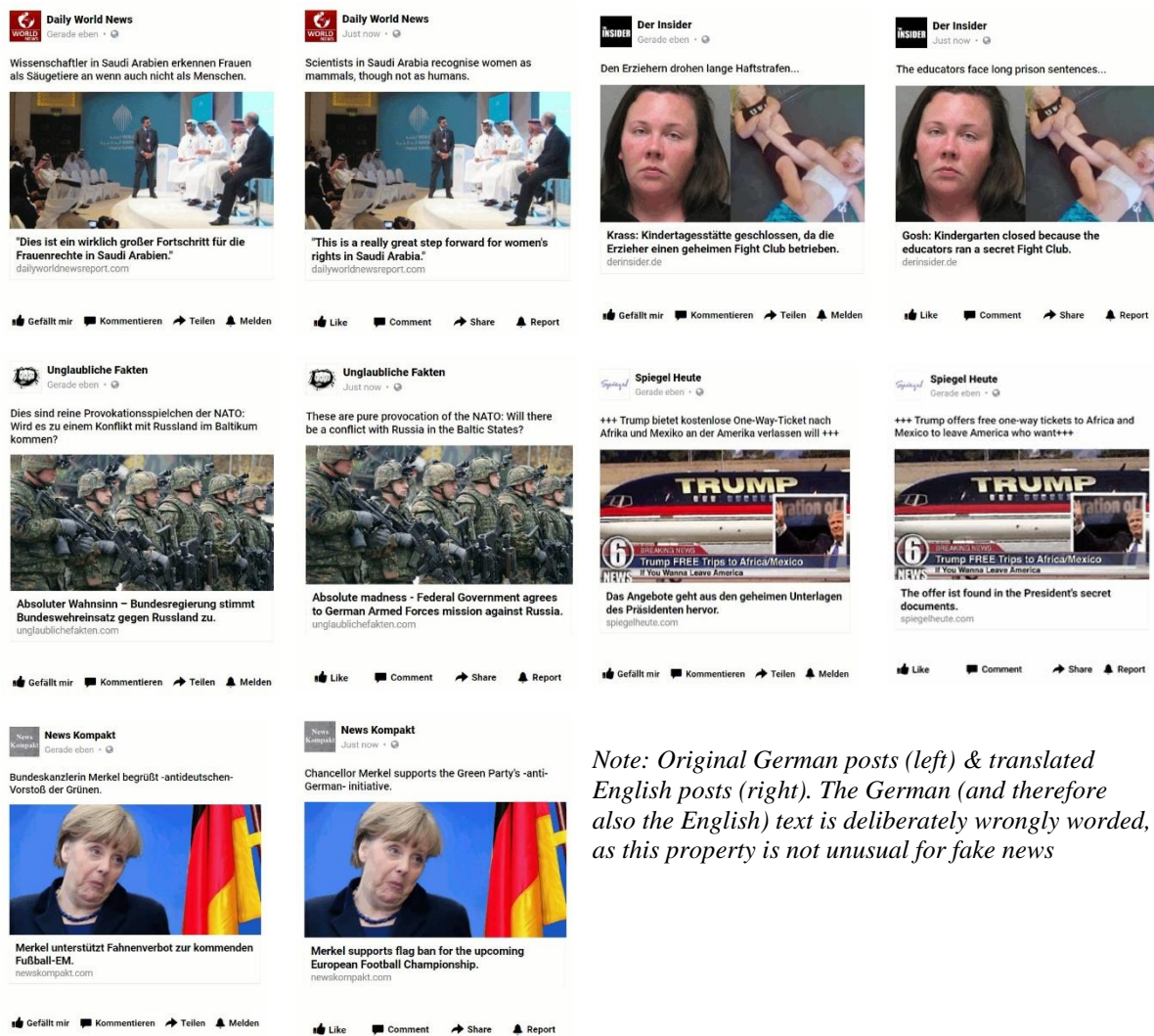
Table 31: Detailed Sample Description of Study 2

Appendix A.11 Stimuli

We focus on a mobile device representation because most users access Facebook via mobile devices (Facebook 2016). Nevertheless, the participants have the possibility to participate in our experiment via desktops or mobile devices. In order to increase comparability, we therefore apply two different views for the news feed. The mobile view shows the news feed about the full size of the browser window. The desktop view first displays the image of a smartphone, which then contains the full screen size news feed identical to the mobile view.











We further extend the news feed by adding injunctive and descriptive SN messages with different wordings.

a) Posts of Study 1 Regarding the Effect of Social Norms



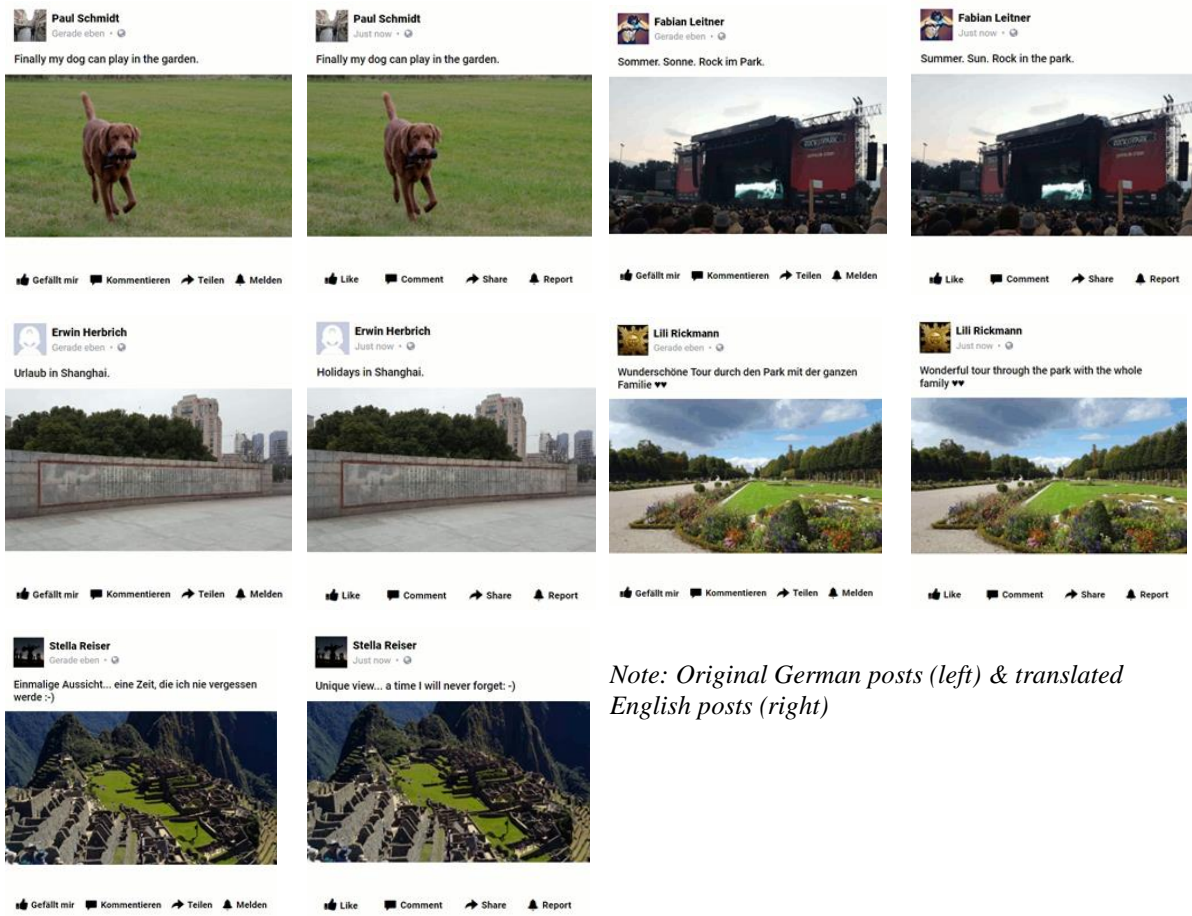
Note: Original German posts (left) & translated English posts (right). The German (and therefore also the English) text is deliberately wrongly worded, as this property is not unusual for fake news

Figure 28: Fake News Posts in Study 1

<p>WELT Gerade eben · 🌐</p> <p>Die Bundespolizei ermittelt und schätzt den Schaden auf mehrere 10.000 Euro.</p>  <p>Kurioser Anblick: In Hamburg haben Unbekannte die Tür einer S-Bahn zugemauert. welt.de</p> <p>👍 Gefällt mir 💬 Kommentieren ➦ Teilen 🚩 Melden</p>	<p>WELT Just now · 🌐</p> <p>The Federal Police estimates the damage to several 10,000 euros.</p>  <p>A curious scene: In Hamburg, unknown people have blocked the door of a suburban railway. welt.de</p> <p>👍 Like 💬 Comment ➦ Share 🚩 Report</p>	<p>FOCUS Online Gerade eben · 🌐</p> <p>Die Detonation war so groß, dass das Dach angehoben wurde.</p>  <p>Am Niederrhein sprengen Täter in der Nacht fast eine ganze Bank in die Luft. focus.de</p> <p>👍 Gefällt mir 💬 Kommentieren ➦ Teilen 🚩 Melden</p>	<p>FOCUS Online Just now · 🌐</p> <p>The detonation was so big that the roof was raised.</p>  <p>At night, perpetrators in the Niederrhein region blow up almost an entire bank. focus.de</p> <p>👍 Like 💬 Comment ➦ Share 🚩 Report</p>
<p>FAZ.NET Gerade eben · 🌐</p> <p>Die drei nordamerikanischen Länder bewerben sich gemeinsam um die WM in neun Jahren.</p>  <p>Amerika, Mexiko und Kanada wollen gemeinsam die Fußball-WM 2026 austragen. faz.net</p> <p>👍 Gefällt mir 💬 Kommentieren ➦ Teilen 🚩 Melden</p>	<p>FAZ.NET Just now · 🌐</p> <p>The three North American countries are applying together for the World Cup in nine years.</p>  <p>The United States, Mexico and Canada want to host the 2026 World Cup together. faz.net</p> <p>👍 Like 💬 Comment ➦ Share 🚩 Report</p>	<p>Tagesschau Gerade eben · 🌐</p> <p>Abgeordnete in Großbritannien debattieren über die Einreise von Präsident Trump.</p>  <p>Mehr als 500.000 Menschen fordern ein Einreiseverbot für den Rechtspopulisten. tagesschau.de</p> <p>👍 Gefällt mir 💬 Kommentieren ➦ Teilen 🚩 Melden</p>	<p>Tagesschau Just now · 🌐</p> <p>Parliamentarians in the UK are debating President Trump's entry.</p>  <p>More than 500,000 people are demanding an entry ban for the right-wing populist. tagesschau.de</p> <p>👍 Like 💬 Comment ➦ Share 🚩 Report</p>
<p>Süddeutsche Zeitung Gerade eben · 🌐</p> <p>Wenn alles wie geplant läuft, wird Kanada im Sommer kommenden Jahres Cannabis legalisieren.</p>  <p>Aufgrund der hohen Nachfrage fürchtet Kanada einen Mangel an Cannabis. sueddeutsche.de</p> <p>👍 Gefällt mir 💬 Kommentieren ➦ Teilen 🚩 Melden</p>	<p>Süddeutsche Zeitung Just now · 🌐</p> <p>If everything goes as planned, Canada will legalise cannabis in the summer of next year.</p>  <p>Due to the high demand, Canada fears a lack of cannabis. sueddeutsche.de</p> <p>👍 Like 💬 Comment ➦ Share 🚩 Report</p>		

Note: Original German posts (left) & translated English posts (right)

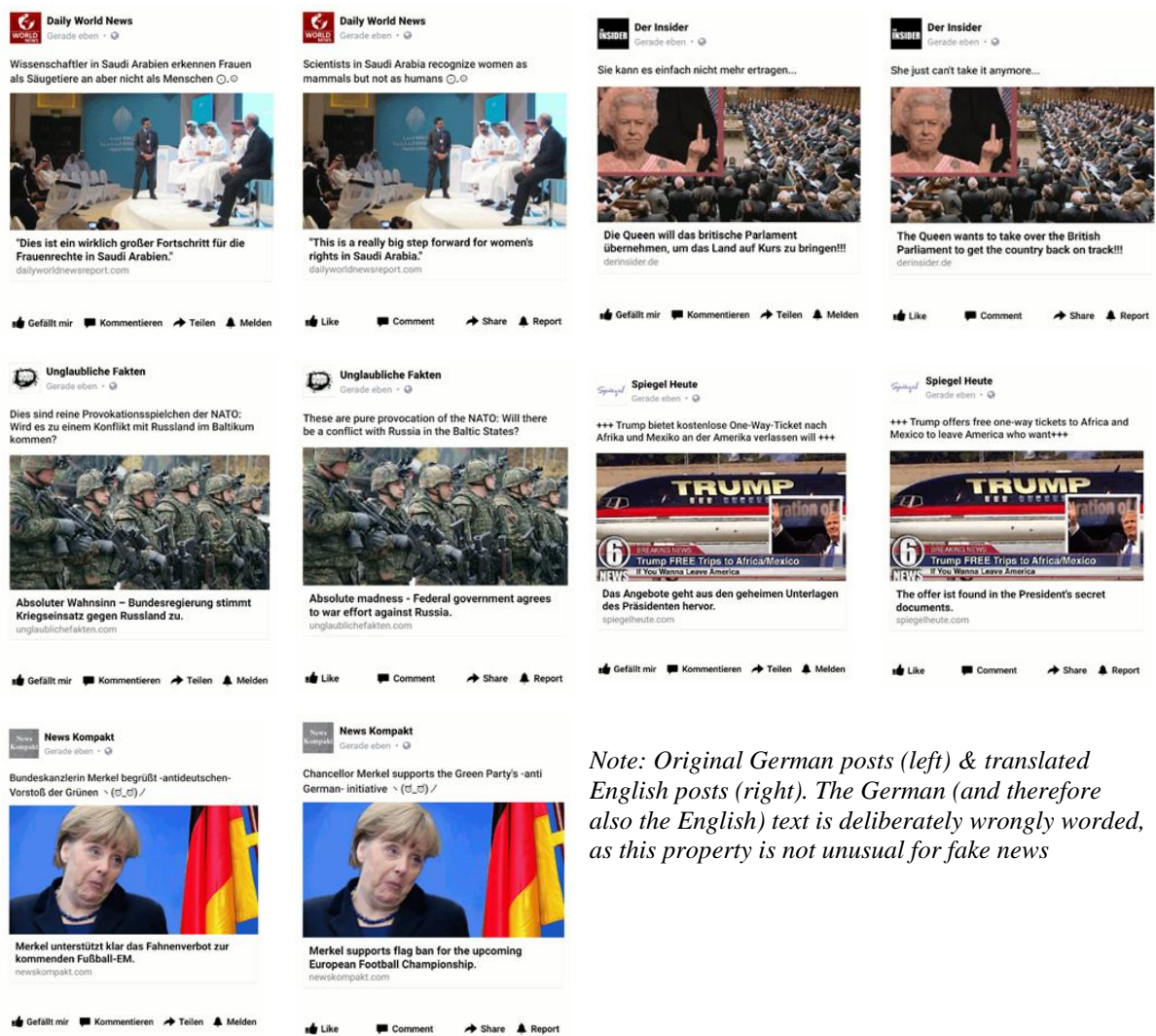
Figure 29: Real News Posts in Study 1



Note: Original German posts (left) & translated English posts (right)

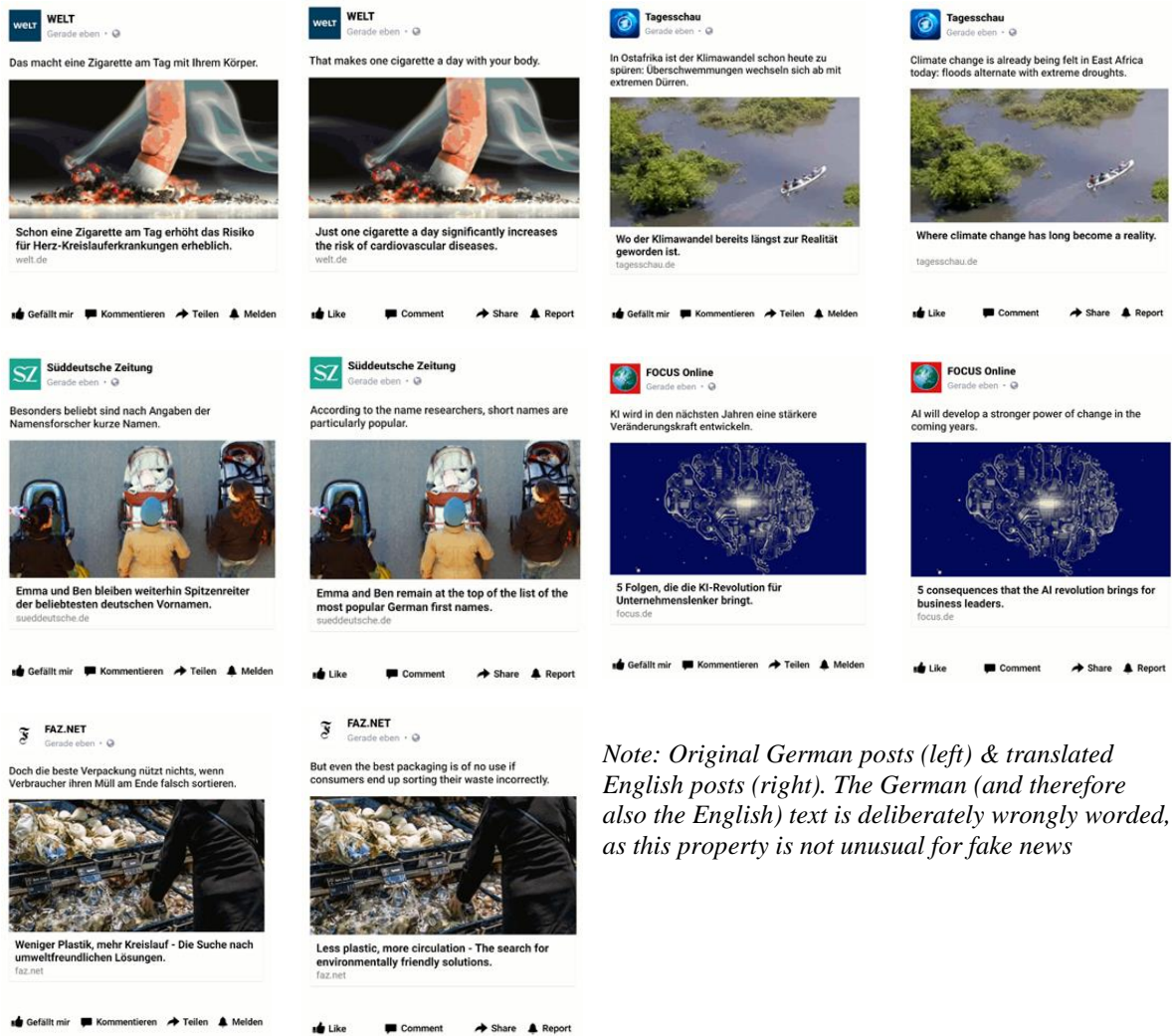
Figure 30: Neutral Posts in Study 1

b) Posts of Study 2 Regarding the Effect of Social Norms



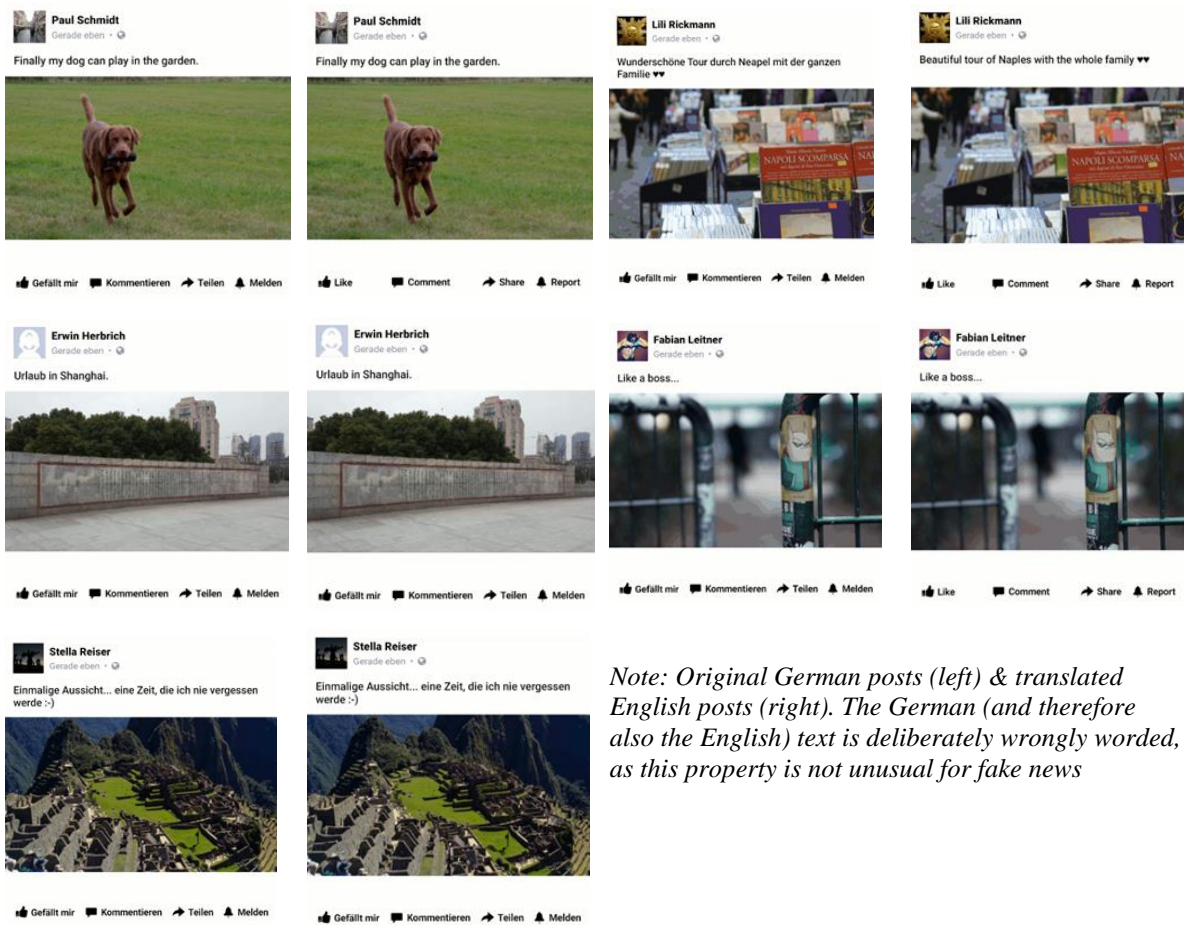
Note: Original German posts (left) & translated English posts (right). The German (and therefore also the English) text is deliberately wrongly worded, as this property is not unusual for fake news

Figure 31: Fake News Posts in Study 2



Note: Original German posts (left) & translated English posts (right). The German (and therefore also the English) text is deliberately wrongly worded, as this property is not unusual for fake news

Figure 32: Real News Posts in Study 2



Note: Original German posts (left) & translated English posts (right). The German (and therefore also the English) text is deliberately wrongly worded, as this property is not unusual for fake news

Figure 33: Neutral Posts in Study 2

Appendix A.12 Results Regarding Structural Equality of Treatments

a) Results of Study 1 Regarding the Effect of Social Norms

Sample property	ANOVA		Levene-Test	
	F-value	p-Value	F-Value	p-Value
Age	1.581	0.194	0.595	0.619
Frequency of social media usage	0.059	0.981	0.059	0.981
Trust in news from social media	0.495	0.686	1.010	0.388

Note: ⁺ $p < 0.1$, * $p < .05$, ** $p < .01$, *** $p < 0.001$

Sample property	χ^2 -test	
	χ^2	p-Value
Gender	4.100	0.231
Occupation	10.262	0.619
Education	23.863	0.470

Note: ⁺ $p < 0.1$, * $p < .05$, ** $p < .01$, *** $p < 0.001$

Table 32: Structural Equality of the Characteristics of the Participants in Study 1

b) Results of Study 2 Regarding the Effect of Social Norms

Sample property	ANOVA		Levene-Test	
	F-value	p-Value	F-Value	p-Value
Age	0.493	0.611	1.875	0.157
Frequency of social media usage	0.100	0.906	0.100	0.906
Trust in news from social media	0.109	0.897	0.145	0.865

Note: ⁺ $p < 0.1$, * $p < .05$, ** $p < .01$, *** $p < 0.001$

Sample property	χ^2 -test	
	χ^2	p-Value
Gender	2.950	0.245
Occupation	6.93	0.838
Education	12.307	0.241

Note: ⁺ $p < 0.1$, * $p < .05$, ** $p < .01$, *** $p < 0.001$

Table 33: Structural Equality of the Characteristics of the Participants in Study 2

Appendix A.13 Development and Validation of the Newsfeed

a) Results of Study 1 Regarding the Effect of Social Norms

Materials: The focus of our work is the reporting of fake news. In order to consider this effect as isolated as possible, we design the fake news to be as conspicuous as possible. Our goal is to make it rather easy for the participants to identify the fake news in order to reduce the effect of uncertainty when identifying fake news. However, we deliberately do not want to tell the participants which news posts are fake news, because such an implementation does not correspond to the real world and also makes the act of reporting unnecessary. A strong limitation of the external validity of our results would be the consequence of clearly identifying fake news as such. Therefore, we modified the fake news posts based on common fake news characteristics such as spelling mistakes, flashy formatting or obviously altered images, to make it easier to identify them. Every fake news story is assigned to a non-existent news source to further reduce authenticity. In addition, the fake news headlines were selected, which have spread in the German-speaking area, since the study took place in Germany.

Besides the fake news, the news feed includes five real news posts. These contain no characteristics of fake news and originate from authentic German news sources. Additionally, headlines were chosen, which have spread in the media in German-speaking countries. However, headlines were selected that cannot be identified as real news at first glance. This also increases the external validity of our results, as the media landscape is not always clear. At the same time, this allows us to observe how the use of SNs affects the reporting behavior of real news.

Overall, especially in comparison to the fake news, the real news can be identified by these characteristics. Besides the news post, the news feed also contains five neutral posts of imaginary friends to create an experimental environment that is as realistic as possible and thus further increase the external validity of the results.

Randomization: The following parts of the experiment were random:

- Assignment of the participants to the treatments
- Display order of the 15 posts by participant
- Selection of the five fake or real news that are marked with a descriptive SN in treatments where the descriptive SN was used (neutral posts were not marked)

- The strength (equivalent to the number of reports) of a descriptive SN to a post in treatments where the descriptive SN was used
- The strength of the descriptive SNs themselves were randomly varied by 10% compared to the base value.

Pre-Test: We conducted a pre-test. The participants of the pre-test ($n = 22$) were recruited via Facebook but attention was paid to use different Facebook groups than for the main study to separate the participants of the pre-test from the participants of the studies. The pre-test suggested feasibility and acceptance of the length of the news feed. Based on the pre-test, we adjusted the fake news stories so that they are more easily identifiable by the participants. Further, we examined and adjusted the strength of the descriptive SNs to ensure that they fit realistically into the news feed. Our first design included a wider interval for the strength of descriptive SNs (1; 10; 100; 1,000; 10,000). However, feedback from the pre-test showed that a norm of 1 is too weak to convince participants that other people report fake news. At the same time, a descriptive SN of 10,000 was found to be too strong. Therefore, we have increased the weakest descriptive SN to 5 and reduced the strongest descriptive SN to 3,125.

Ensuring Response Quality: To ensure the manipulation validity of our studies, it was ensured for each participant that the study had been fully processed. 349 participants have completed the survey. To further improve the quality of our data, we have taken further measures. First, tracking points were implemented into the newsfeed to identify whether a participant has seen all posts in the newsfeed. Only participants who had read the newsfeed completely were considered. Second, we have removed all participants who use social media less than once a week. Our final sample for survey 1 contains 320 participants.

Experiment Validation: To validate our approach, we consider the results of the fake assessment run. Although the fake assessment run should be used with caution (due to the repeated presentation), it is suitable as an indicator to identify which posts the participants considered as fake. Table 34 summarizes the number of reports for the posts in the fake detection run. The results show that on average 79% of the participants identified fake news as such. In contrast, only 20% of the participants reported real news as fake news.

	News Content	Number of reports	Fake detection rate
Fake News	Merkel Flag Ban	273	85.31%
	NATO Mission	262	81.88%
	One-way Ticket	259	80.94%
	Woman Rights	250	78.13%
	Kindergarten Fight Club	219	68.44%
Real News	Legal Cannabis	94	29.38%
	Host World Cup	94	29.38%
	Trump Entry Ban	57	17.81%
	Blocked Waggon	50	15.63%
	Blown Up bank	25	7.81%

Table 34: Number of Reports per Post in the Fake Assessment Run of Study 1

The following Table 35 compares fake news and real news by counting how often the fake and real news were reported and how often they were not reported. The Fisher's exact test shows a highly significant empirical difference (p-value < 0.001). These results are in line with our expectations, since both the fake news and the real news were identified as such by the participants.

	Reported	Not reported
Fake News	1,263	337
Real News	320	1,280

Table 35: Overall Number of Reports in the Fake Assessment Run of Study 1

b) Results of Study 2 Regarding the Effect of Social Norms

Materials: For Study 2 we used almost the same materials as for Study 1. However, we modified four of the five original fake news to further decrease the level of authenticity. Emoticons (such as ‘\ (ಠ_ಠ)ノ’ and ‘ಠ.ಠ’) were added to the news posts about women's rights in Saudi Arabia and the ban on flags at the European Championship to give participants the additional feeling that the news source is not neutral, but that the author wants to give a special feeling and opinion, as it is often the case with fake news. At the same time, we changed the term 'German Armed Forces Mission' to 'War mission' in the fake news post about the NATO mission. We intended to differentiate the statement even more from the real world and thus make it less credible.

Additionally, we replaced one fake news completely (“Kindergarten Fight Club”), because it was recognized much less by the participants than the others. We assume that this fake news was harder to identify because it does not relate to events with which the participants frequently come into contact.

The real news was also revised. Since the real news stories were outdated at the time of Study 2, we replaced them with more recent ones. We have also adjusted the neutral posts, as these explicitly referred to events in the summer (the time of Study 1). Since Study 2 took place in winter, we have exchanged these posts to make our news feed more realistic.

Randomization: As in Study 1.

Pre-Test: Since our material and procedure had already been validated by the pre-test and Study 1, we did not conduct another pre-test for Study 2.

Ensuring Response Quality: We have taken the measures described above for Study 2 with two additions. First, we asked the participants whether they had already participated in Study 1 and removed all participants who confirmed this. Second, we added a control item to each of our scales to record the motivational factors for reporting. This control item corresponds to the inverted statement of another, which is why a response at the same end of the scale is inconsistent. Table 36 presents the items described.

Target construct	Test item	Control item
Motivation to report fake news	I would like to avoid negative consequences that arise from the distribution of fake news	I am not interested in the negative consequences that arise from the distribution of fake news
Motivation to not report fake news	I've never seen fake news	I often come in contact with fake News

Table 36: Summary of Control Items in Study 2

We have removed all participants who answered the control item inconsistently. Therefore, our final data set for Study 2 consists of 157 participants.

Experiment Validation: To validate our approach, we consider the results of the updated fake assessment run. Table 37 summarizes the number of reports for the posts in the fake detection run. The results show that on average 87% of the participants identified fake news as such. In contrast, only 3% of the participants reported real news as fake news.

	News Content	Rated as fake	Fake detection rate
Fake News	British Parliament	153	97.45%
	Merkel Flag Ban	143	91.08%
	NATO Mission	140	89.17%
	Woman Rights	127	80.89%
	One-way Ticket	118	75.16%
Real News	AI Influence	7	4.46%
	Climate Chang	7	4.46%
	Effects of Smoking	4	2.55%
	Reuse of Shopping Bags	3	1.91%
	First Name Raking	3	1.91%

Table 37: Number of News Stories Rated as Fake News in Study 2.

Table 38 compares fake news and real news by counting how often the fake and real news were rated as fake and how often they were not rated as fake. The Fisher's exact test shows a highly significant empirical difference ($p\text{-value} < 0.001$). These results are in line with our expectations, since we deliberately made the difference between fake and real news more extreme for Study 2.

	Rated as fake	Not rated as fake
Fake News	681	104
Real News	24	761

Table 38: Overall Number of News Stories Rated as Fake News in Study 2

Appendix A.14 List of Scales

a) List of Scales for Study 1 Regarding the Effect of Social Norms

Construct	Question	Scale type	Response options	Source
Social media as a source of news	How often do you use social media to obtain information on current events, public affairs and political issues?	9-point (ordinal) Likert scale	Never; Less than once a week; Once a week; Twice a week; Three times a week; Four times a week; Five times a week; Six times a week; Daily	Based on the scale for measuring the frequency of the use of different media as a news source by Gil de Zúñiga et al. (2012).
	How much trust do you have in information about current events, public affairs, political issues that comes from social media?	6-point (ordinal) Likert scale	No trust at all; Little trust; Rather little trust; Rather a lot of trust; A lot of trust; A great deal of trust	
Indicators for Fake News	What features did you use in the survey and do you generally use to identify a post as fake news?	Multiple Choice with optional text input	Source/Creator; Content of the text; Spelling mistake in the text; Striking formatting of the text; Picture in the post; Number of times a mail has already been reported; others	Self-developed
Usage of social media	How often do you use social media in your everyday life?	6-point (ordinal) Likert scale	Never; Less than once a week; Once a week; Several times a week; Once a day; Several times a day	Self-developed
	How much content do you generate in social media (e.g. by posting, liking, commenting, sharing, etc.) compared to people from your personal environment?	7-point (ordinal) Likert scale	No content at all; Very little content; Little content; Rather little content; Rather a lot of content; A lot of content; A great deal of content	
	How many posts do you like per day?	Numerical input	Integer value equal or greater than 0	
	How many posts do you comment per day?	Numerical input	Integer value equal or greater than 0	
	How many posts do you share per day?	Numerical input	Integer value equal or greater than 0	
Automatic vs. mindful thought processing	I don't like to have to do a lot of thinking.	5-level semantic differential	Min: Strongly disagree Max: Strongly agree	Self-translated scale to measure the degree to which one engages in mindful versus automatic thought-processing of news by Maksl et al. (2015).
	I try to avoid situations that require thinking in depth about something.	5-level semantic differential	Min: Strongly disagree Max: Strongly agree	
	I prefer to do something that challenges my thinking abilities rather than something that requires little thought.	5-level semantic differential	Min: Strongly disagree Max: Strongly agree	
	I prefer complex to simple problems.	5-level semantic differential	Min: Strongly disagree Max: Strongly agree	
	Thinking hard and for a long time about something gives me little satisfaction.	5-level semantic differential	Min: Strongly disagree Max: Strongly agree	

Construct	Question	Scale type	Response options	Source
Cognitive reflection test	A racket and a ball cost a total of 1.10 Euro. The racket costs 1.00 Euro more than the ball. How much does the ball cost?	Numerical input	Decimal number equal or greater than 0	Self-translated scale to measure reflective reasoning by Frederick (2005).
	Five machines take 5 minutes to produce five products. How many minutes does it take 100 machines to produce 100 products?	Numerical input	Integer value equal or greater than 0	
	Lilies grow in a lake. Every day, the area that covers the lilies on the lake doubles. If it takes 48 days until the lilies cover the whole lake, how long would it take until the lilies cover half of the lake?	Numerical input	Integer value equal or greater than 0	
Gender	What is your gender?	Single Choice	female; male	Based on the standard demographic questions of the Sosci Survey tool (www.soscisurvey.de)
Age	How old are you?	Numerical input	Integer value equal or greater than 0	
Educational level	What is your educational level? Please select the highest level of education you have achieved so far.	Single Choice with optional text input	Leaving school without graduation; Still in school; junior high school (original: "Volks-, Hauptschulabschluss", "Mittlere Reife", "Realschule") or equivalent degree; Completed apprenticeship; Senior High School (Original: "(Fach-)Abitur"); Bachelor's degree; Master's degree; Diploma; Doctorate / postdoctoral qualification; Different degree [free text option]	
Occupation	Which of the following categories best describes your employment status?	Single Choice with optional text input	Pupil; In an apprenticeship; Student; Salaried employee; Civil servant; Independent; Unemployed; Others [free text option]	

Table 39: Description of all Scales Used in Study 1

b) List of Scales for Study 2 Regarding the Effect of Social Norms

Construct	Question	Scale type	Response options	Source
Social media as a source of news	How often do you use social media to obtain information on current events, public affairs and political issues?	9-point (ordinal) Likert scale	Never; Less than once a week; Once a week; Twice a week; Three times a week; Four times a week; Five times a week; Six times a week; Daily	Based on the scale for measuring the frequency of the use of different media as a news source by Gil de Zúñiga et al. (2012).
	How much trust do you have in information about current events, public affairs, political issues that comes from social media?	6-point (ordinal) Likert scale	No trust at all; Little trust; Rather little trust; Rather a lot of trust; A lot of trust; A great deal of trust	
Assessment of news posts	Please rate whether the post shown above is a true news story or fake news:	5-point (ordinal) Likert scale	True news story; Rather a true news story; I can't tell; Rather fake news; Fake news	Self-developed
Indicators for Fake News	What features did you use in the survey and do you generally use to identify a post as fake news?	Multiple Choice with optional text input	Source/Creator; Content of the text; Spelling mistake in the text; Striking formatting of the text; Picture in the post; Number of times a mail has already been reported; others	Self-developed
Reasons to report fake news	I would like to improve the living conditions for myself and other people	5-level semantic differential	Min: Strongly disagree Max: Strongly agree	Self-developed
	It is important to me to help other people to form their opinion based on true facts	5-level semantic differential	Min: Strongly disagree Max: Strongly agree	
	I would like to avoid negative consequences that result from the dissemination of fake news	5-level semantic differential	Min: Strongly disagree Max: Strongly agree	
	Fake News cause an unpleasant feeling in me	5-level semantic differential	Min: Strongly disagree Max: Strongly agree	
	I would like to improve the quality of the social media platform I use	5-level semantic differential	Min: Strongly disagree Max: Strongly agree	
	I expect a material or financial incentive	5-level semantic differential	Min: Strongly disagree Max: Strongly agree	
	A correct news landscape is important to me	5-level semantic differential	Min: Strongly disagree Max: Strongly agree	
	I want to share my knowledge with others	5-level semantic differential	Min: Strongly disagree Max: Strongly agree	
	I hope for appreciation from my social environment	5-level semantic differential	Min: Strongly disagree Max: Strongly agree	
	I am not interested in the negative consequences that arise from the distribution of fake news	5-level semantic differential	Min: Strongly disagree Max: Strongly agree	
I do not want that the relevance of the topic will be reduced by the dissemination of fake news	5-level semantic differential	Min: Strongly disagree Max: Strongly agree		

Appendix A - Basic User Assistance Systems to Combat Fake News

Construct	Question	Scale type	Response options	Source
Hurdles to report fake news	I am not interested in the public opinion	5-level semantic differential	Min: Strongly disagree Max: Strongly agree	
	The procedure of reporting is too complex for me	5-level semantic differential	Min: Strongly disagree Max: Strongly agree	
	I've never seen fake news	5-level semantic differential	Min: Strongly disagree Max: Strongly agree	
	I do not believe that the act of reporting has any effect or counteracts the dissemination of fake news	5-level semantic differential	Min: Strongly disagree Max: Strongly agree	
	I don't know how to report fake news	5-level semantic differential	Min: Strongly disagree Max: Strongly agree	
	Reporting fake news offers me no material or financial incentive	5-level semantic differential	Min: Strongly disagree Max: Strongly agree	
	Fake news is subject to the freedom of speech	5-level semantic differential	Min: Strongly disagree Max: Strongly agree	
	Fake news serves for entertainment	5-level semantic differential	Min: Strongly disagree Max: Strongly agree	
	I often come in contact with fake news	5-level semantic differential	Min: Strongly disagree Max: Strongly agree	
	I am only a consumer of news and do not actively participate in the public discourse	5-level semantic differential	Min: Strongly disagree Max: Strongly agree	
Usage of social media	How often do you use social media in your everyday life?	6-point (ordinal) Likert scale	Never; Less than once a week; Once a week; Several times a week; Once a day; Several times a day	Self-developed
	How much content do you generate in social media (e.g. by posting, liking, commenting, sharing, etc.) compared to people from your personal environment?	7-point (ordinal) Likert scale	No content at all; Very little content; Little content; Rather little content; Rather a lot of content; A lot of content; A great deal of content	
	How many posts have you liked in the last month compared to people from your personal environment?	7-point (ordinal) Likert scale	None at all; Very few; Few; Rather few; Rather many; Many; A great many	
	How many posts have you commented on in the last month compared to people from your personal environment?	7-point (ordinal) Likert scale	None at all; Very few; Few; Rather few; Rather many; Many; A great many	
	How many posts have you shared in the last month compared to people from your personal environment?	7-point (ordinal) Likert scale	None at all; Very few; Few; Rather few; Rather many; Many; A great many	
	How many posts have you reported in the last month compared to people from your personal environment?	7-point (ordinal) Likert scale	None at all; Very few; Few; Rather few; Rather many; Many; A great many	

Construct	Question	Scale type	Response options	Source
Gender	What is your gender?	Single Choice	female; male	Based on the standard demographic questions of the Sosci Survey tool (www.soscisurvey.de)
Age	How old are you?	Numerical input	Integer value equal or greater than 0	
Educational level	What is your educational level? Please select the highest level of education you have achieved so far.	Single Choice with optional text input	Leaving school without graduation; Still in school; junior high school (original: "Volks-, Hauptschulabschluss", "Mittlere Reife", "Realschule") or equivalent degree; Completed apprenticeship; Senior High School (Original: "(Fach-)Abitur"); Bachelor's degree; Master's degree; Diploma; Doctorate / postdoctoral qualification; Different degree [free text option]	
Occupation	Which of the following categories best describes your employment status?	Single Choice with optional text input	Pupil; In an apprenticeship; Student; Salaried employee; Civil servant; Independent; Unemployed; Others [free text option]	

Table 40: Description of All Scales Used in Study 2

Appendix A.15 Results of the Brant test for Ordered Logistic Regression

a) Results of Study 1 Regarding the Effect of Social Norms

Base Treatment	Treatment	Coefficient	df	Probability
Control	Omnibus	4.205	12	0.979
	Injunctive SN	0.301	4	0.990
	Descriptive SN	1.198	4	0.878
	Combined	2.970	4	0.563
Combined	Omnibus	4.205	12	0.979
	Control	2.970	4	0.563
	Injunctive SN	2.126	4	0.713
	Descriptive SN	0.924	4	0.921

Note: Brant test to test the assumption of proportional odds for the ordered logistic regression models of Study 1.

+ $p < 0.1$, * $p < .05$, ** $p < .01$, *** $p < 0.001$

Table 41: Results of the Brant Test in Study 1

b) Results of Study 2 Regarding the Effect of Social Norms

Base Treatment	Treatment	Coefficient	df	Probability
Control	Omnibus	9.956	8	0.268
	Positive descriptive SN	2.253	4	0.689
	Negative descriptive SN	5.450	4	0.244

Note: Brant test to test the assumption of proportional odds for the ordered logistic regression models of Study 2.

+ $p < 0.1$, * $p < .05$, ** $p < .01$, *** $p < 0.001$

Table 42: Results of the Brant Test in Study 2

Appendix A.16 Results of Real News Reporting

Base Treatment	Treatment	Coefficient	Standard Error	p-Value	Odds Ratio
Control	Injunctive SN	0.847	0.709	0.232	2.333
	Descriptive SN	1.200	0.687	0.080 +	3.320
	Combined	1.482	0.666	0.026 *	4.402
	Nagelkerke's R ² : 0.037				
Combined	Control	-1.482	0.666	0.026 *	0.227
	Injunctive SN	-0.635	0.503	0.207	0.530
	Descriptive SN	-0.281	0.472	0.552	0.755
	Nagelkerke's R ² : 0.037				

Note: Ordered logistic regression to compare the four treatments with respect to real news reporting. In the first model the control treatment is the baseline, whereas in the second model the combined treatment is the baseline.

*+ p < 0.1, * p < .05, ** p < .01, *** p < 0.001*

Table 43: Results of Real News Reporting of Study 1

Appendix A.17 Results of the Ordered Logistic Regression Including Awareness

The introduction to the news feed is an awareness training that aims to teach the participants that they can report content and how to do so without telling them that they should report. Because every participant goes through the identical introduction, regardless of which treatment they are in, we assume that awareness is equally present in each of the treatments. However, to control for the possible effects of different degrees of awareness, we distinguish two levels of awareness, depending on whether the participants had interacted with the report button in the interactive tutorial or not. If a person has used the report feature in the tutorial, we assume a high degree of awareness. If a person has not used the report button during the tutorial, we use this as indication of a lower level of awareness of the reporting feature. We extend our ordinal logistic regression analysis to include awareness (high and low).

a) Results of Study 1 Regarding the Effect of Social Norms

Base Treatment		Coefficient	Standard Error	p-value	Odds Ratio (Effect Size)	Related Hypothesis	
Control	Treatments	Injunctive SN	0.673	0.326	0.039 *	1.960 (small)	H 3.3.1
		Descriptive SN	0.456	0.335	0.173	1.578 (small)	H 3.3.2
		Combined	1.374	0.321	< 0.001 ***	3.951 (medium)	H 3.3.3
	Control variable	Awareness	-0.086	0.245	0.724	0.918 (less than small)	
	Intercepts	0 1	0.864	0.305	< 0.005 **		
		1 2	1.730	0.316	< 0.001 ***		
		2 3	2.250	0.328	< 0.001 ***		
		3 4	3.343	0.372	< 0.001 ***		
		4 5	4.495	0.484	< 0.001 ***		
	Nagelkerke's R ² : 0.068						

Note: Ordered logistic regression to compare the three treatments with respect to fake news reporting. The control treatment is the baseline.

+ $p < 0.1$, * $p < .05$, ** $p < .01$, *** $p < 0.001$

Table 44: Results of Fake News Reporting While Controlling For Awareness

b) Results of Study 2 Regarding the Effect of Social Norms

Base Treatment			Coefficient	Standard Error	p-value	Odds Ratio (Effect Size)	Related Hypothesis
Control	Treatments	Positive descriptive SN	0.333	0.370	0.368	1.395 (less than small)	H 3.3.2
		Negative descriptive SN	0.110	0.396	0.781	1.116 (less than small)	
	Control variable	Awareness	0.490	0.311	0.116	1.632 (small)	
	Intercepts	0 1	0.604	0.333	0.070 ⁺		
		1 2	1.318	0.347	< 0.001 ***		
		2 3	1.939	0.365	< 0.001 ***		
		3 4	2.279	0.380	< 0.001 ***		
		4 5	2.901	0.425	< 0.001 ***		
	Nagelkerke's R ² : 0.022						

Note: ordered logistic regression to compare the three treatments with respect to fake news reporting. The control treatment is the baseline.

⁺ $p < 0.1$, * $p < .05$, ** $p < .01$, *** $p < 0.001$

Table 45: Results of Fake News Reporting While Controlling for Awareness

Appendix A.18 Effects of Strength of Descriptive Social Norms

Natural interaction run – fake news posts							
Fake posts		Coefficients					
		0	5	25	125	625	3125
Baselines	0	-1.473 ***	0.723 *	0.823 *	1.160 ***	1.306 ***	0.570
	5		-0.750	0.100	0.437	0.583	-0.153
	25			-0.651 +	0.337	0.484	-0.252
	125				-0.314	0.147	-0.589 +
	625					-0.167	-0.736 *
	3125						-0.903 **

Note: Coefficients (upper triangle) and intercepts (diagonal) of the logistic regression models with different baselines for fake news posts in the natural interaction run.

+ $p < 0.1$, * $p < .05$, ** $p < .01$, *** $p < 0.001$

Table 46: Effects of Strength of Descriptive Social Norms on Fake News in Study 1

Natural interaction run – real news posts							
Fake posts		Coefficients					
		0	5	25	125	625	3125
Baselines	0	-3.475 ***	0.142	0.361	1.047	0.702	0.836
	5		-3.332 **	0.219	0.905	0.560	0.693
	25			-3.114 ***	0.686	0.341	0.475
	125				-2.428 ***	-0.345	-0.211
	625					-2.772 ***	0.134
	3125						-2.229 ***

Note: Coefficients (upper triangle) and intercepts (diagonal) of the logistic regression models with different baselines for real news posts in the natural interaction run.

+ $p < 0.1$, * $p < .05$, ** $p < .01$, *** $p < 0.001$

Table 47: Effects of Strength of Descriptive Social Norms on Real News in Study 1

Appendix B Interactive User Assistance Systems to Promote Behavior Change

Appendix B.1 Overview of all Reviewed Papers

Authors(s) (Year)	Publication	Publishing category	DHR design (social role, physical, psychological, language)	Psychological constructs	Intervention types	Targeted behavior	Experiment information
Bickmore et al. (2005a)	<i>Interacting with Computers</i>	J, Q3 (HCI)	EA: exercise advisor 2D, dynamic, upper body mimics, speech synthetic voice	Satisfaction, repetitiveness, friendliness, trust, informativeness, liking, ease of use, relationship (empathy)	EDU, PERS, ENAB, ENVR, INC	PA (*, 2m, pedometer, intention to use)	Field (21)
Bickmore et al. (2005b)	<i>Patient Education and Counseling</i>	J, Q1 (Medicine)	EA: exercise advisor 2D, dynamic, face eyebrow raises, gaze, posture shifts, nods (relational vs. non-relational) synthetic speech	Liking, relatedness, usefulness, caring, honesty	EDU, PERS, ENAB	PA (*, 1w, pedometer, intention to use)	Field (91)
de Rosi et al. (2006)	<i>Journal of Biomedical Informatics</i>	J, Q1 (Health Informatics)	EA: dietary expert [WOz] 3D, dynamic, head mimics, speech synthetic voice	Credibility, plausibility, clarity, usefulness, persuasiveness, sincerity, likeability, naturality, intelligence, competence	PERS	Nutrition	Lab (30)
van Vugt et al. (2006)	<i>Intelligent Virtual Agents</i>	C (B-rank)	EA: health advisor 2D, static, full-body - text bubble	Similarity, valence, distance, aesthetics, realism, ethics (trustworthiness & credibility)	EDU	PA, Nutrition (intention to use)	Survey (278)
Bickmore et al. (2007)	<i>Persuasive Technology</i>	C (B-rank)	EA: health advisor 2D, dynamic, face only gaze, lips, eyebrows, nods, posture text bubble	Politeness, annoying	EDU, PERS, ENVR	PA (*, instant rest time, intention to use)	Lab (29)
Fox et al. (2009)	<i>Presence: Teleoperators and Virtual Environments</i>	J, Q3 (HCI)	AV: virtual self 3D, dynamic, full body - -	Presence	PERS, MOD	Nutrition (*, instant food choice)	Lab (69)
Fox and Bailenson (2009)	<i>Media Psychology</i>	J, Q1 (Applied Psychology)	AV: virtual self / other 3D, dynamic, full-body - -	Resemblance	COE, INC, MOD, TRA	PA (*, 1d, self-reported activity)	Lab (63) Lab (60) Lab (75)
Jin (2009)	<i>CyberPsychology & Behavior</i>	J, Q1 (Applied Psychology)	AV: ideal/actual self 3D, dynamic, full-body gestures, mimics -	Interactivity, immersion, attractiveness	PERS	Nutrition	Lab (126)
Mazzotta et al. (2009)	<i>Intelligent Virtual Agents</i>	C (B-rank)	EA: dietary expert 3D, dynamic, head mimics, speech synthetic voice	Satisfaction, helpfulness, easiness, persuasiveness, reliability, validity	PERS	Nutrition	Lab (60)

Appendix B - Interactive User Assistance Systems to Promote Behavior Change

Authors(s) (Year)	Publication	Publishing category	DHR design (social role, physical, psychological, language)	Psychological constructs	Intervention types	Targeted behavior	Experiment information
Peng (2009)	<i>Health Communication</i>	J, Q1 (Health)	EA+AV: virtual self & health-conscious college students, a school dietitian, a personal trainer at the gym, and cafeteria workers 2D, static, face only - text box	Enjoyment, credibility	EDU, TRA, MOD	PA, Nutrition (efficacy)	Lab (40)
Schulman and Bickmore (2009)	<i>Proceedings of the 4th International Conference on Persuasive Technology</i>	C (B-rank)	EA: health counselor 2D, dynamic, full-body gestures, eye movement, facial emotion synthesized speech, text	Persuasiveness, competence, honesty, trustworthiness, boldness	EDU, PERS, ENAB	PA	Lab (47)
van Vugt et al. (2009)	<i>International Journal of Human Computer Studies</i>	J, Q1 (HCI)	EA: health advisor 2D, static, full-body - text bubble	Involvement, interpersonal distance, perceived similarity, perceived ethics (trustworthiness & credibility)	EDU	PA, Nutrition (intention to use)	Survey (80) Survey (259)
Bickmore et al. (2010)	<i>Applied Artificial Intelligence</i>	J, Q3 (AI)	EA: exercise counselor 2D, dynamic, upper body gestures, mimics MC answers, pedometer	Repetitiveness, enjoyment, honesty	EDU, PERS, ENAB	PA (*, 188- 267d, 5- 37d, pedometer, intention to use)	Field (24) Field (26)
Creed and Beale (2012)	<i>Interacting with Computers</i>	J, Q3 (HCI)	EA: nutrition coach 3D, dynamic, face mimics, speech recorded voice	Likeability, caring, trustworthiness, perceived intelligence, positivity, supportedness	EDU, PERS, ENAB, ENVR	Nutrition	Survey (50)
Johnston et al. (2012)	<i>Proceedings of the 45th Annual Hawaii International Conference on System Sciences</i>	C (A-rank)	EA+AV: ideal/actual self in group intervention with agents 3D, dynamic, full- body ? text	-	EDU, TRA, PERS, ENAB, MOD, ENVR	PA, Nutrition (*, 12w, weight loss compared to f2f interv.)	Field (54)
Kim and Sundar (2012b)	<i>Computers in Human Behavior</i>	J, Q1 (HCI)	EA+AV: CDC agent, ideal/actual self 3D, dynamic, full body gestures text field, MC answers	Attractiveness, vividness of body perception (presence), perceived risk	EDU, PERS, MOD	PA, Nutrition, smoking, alcohol (*, instant coupon choice)	Lab (69)
Watson et al. (2012)	<i>Journal of Medical Internet Research</i>	J, Q1 (Health Informatics)	EA: PA coach 2D, dynamic, full-body gestures, gaze synthetic speech	Satisfaction	EDU, TRA, PERS, ENAB	PA (*, 12w, pedometer)	Field (70)
Bickmore et al. (2013a)	<i>Patient Education and Counseling</i>	J, Q1 (Medicine)	EA: health counselor 2D, dynamic, upper body gestures, gaze synthetic speech	Satisfaction, ease of use	EDU, PERS, ENAB	PA, Nutrition (*, 2m, pedometer, self-report, intention to use)	Field (122)

Appendix B - Interactive User Assistance Systems to Promote Behavior Change

Authors(s) (Year)	Publication	Publishing category	DHR design (social role, physical, psychological, language)	Psychological constructs	Intervention types	Targeted behavior	Experiment information
Bickmore et al. (2013b)	<i>Journal of the American Geriatrics Society</i>	J, Q1 (Geriatrics)	EA: exercise coach 2D, dynamic, full body gestures, facial emotions synthetic speech	Satisfaction, helpfulness	EDU, PERS, ENAB	PA (*, 2m, 12m, pedometer)	Field (263)
King et al. (2013)	<i>Journal of Health Communication</i>	J, Q1 (Health)	EA: virtual PA advisor 2D, dynamic, upper body gestures, facial emotion synthetic speech	Caring, social distance, credibility, ease of use	EDU, PERS, ENAB	PA (*, 4m, pedometer, intention to use)	Field (40)
Klaassen et al. (2013a)	<i>Journal on Multimodal User Interfaces</i>	J, Q2 (HCI)	EA: health professional 2D, static, upper body gestures & body animations synthesized voice, text	Pragmatic quality, hedonic quality, attractiveness	PERS, ENAB	PA	Field (9)
Klaassen et al. (2013b)	<i>ACM International Conference Proceeding Series, International Conference on Pervasive Technologies Related to Assistive Environments</i>	C (unranked)	EA: health professional 2D, static, upper body gestures & body animations synthesized voice, text	Enjoyment, credibility, naturalness, enthusiasm, glanceability	PERS, ENAB	PA	Field (14)
Lisetti et al. (2013)	<i>ACM Transactions on Management Information Systems</i>	J, Q1 (CS)	EA: alcohol counselor 3D, dynamic, upper body gestures, mimics (empathic vs non-empathic vs text) synthetic voice	Empathy, anthropomorphism, animacy, likeability, intelligence, trust, social presence, usefulness, enjoyment, ease of use, sociability, anxiety, social influence, safety	EDU, PERS, ENVR, ENAB	Alcohol (intention to use)	Lab (81)
Morie et al. (2013)	<i>Distributed, Ambient, and Pervasive Interactions</i>	C (unranked)	AV: virtual self 3D, dynamic, full-body - -	Similarity	MOD	PA (*, 1d, self-report)	Survey (143)
Napolitano et al. (2013)	<i>Journal of Diabetes Science and Technology</i>	J, Q1 (Bio-engineering)	AV: actual self 3D, dynamic, full-body - recorded voice (instructor voice)	Satisfaction	EDU, PERS, MOD	PA, Nutrition (*, 1m, confidence, self-efficacy, weight loss)	Field (8)
Song et al. (2013)	<i>Computers in Human Behavior</i>	J, Q1 (HCI)	AV: actual/future self 2D, static, full body with photo face - -	Identification, perceived susceptibility	MOD, COE, PERS	Smoking (intention to quit)	Lab (62)
Ahn et al. (2014b)	<i>Intelligent Virtual Agents</i>	C (B-rank)	AV: virtual self / other 3D, dynamic, full body - -	Personal relevance, self-presence	MOD, COE	Nutrition (risk perception)	Lab (47)
Friederichs et al. (2014)	<i>Journal of Medical Internet Research</i>	J, Q1 (Health Informatics)	EA: coach 2D, dynamic, upper body - text bubble	Personal relevance, trustworthiness, appreciation	PERS, ENAB	PA (*, 1m, self-report)	Field (958)
Kim et al. (2014)	<i>Computers in Human Behavior</i>	J, Q1 (HCI)	AV: virtual self 3D, dynamic, full body -	Presence, enjoyment	TRA	PA (*, heart rate, intention to change)	Lab (119)

Appendix B - Interactive User Assistance Systems to Promote Behavior Change

Authors(s) (Year)	Publication	Publishing category	DHR design (social role, physical, psychological, language)	Psychological constructs	Intervention types	Targeted behavior	Experiment information
Li et al. (2014)	<i>Games for Health Journal</i>	J, Q1 (Health)	AV: virtual self 3D, dynamic, full body ? -	-	TRA, PERS	PA (intention to exercise)	Lab (140)
Peña and Kim (2014)	<i>Computers in Human Behavior</i>	J, Q1 (HCI)	AV: virtual self & virtual other 3D, dynamic, full body - -	-	TRA, PERS	PA (*, activity sensors)	Lab (94)
Schmeil and Suggs (2014)	<i>Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)</i>	J, Q2 (Computer Science)	AV: actual/future self 3D, dynamic, full body - -	Perc. healthiness	PERS, MOD	PA, Nutrition (motivation, intention to change)	Survey (512)
Vainio et al. (2014)	<i>Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare</i>	C (unranked)	EA: supporter 2D, static, face only mimics speech bubble	-	TRA, PERS, ENAB	PA, Nutrition (*, 1m, self- report)	Field (66)
Yasavur et al. (2014)	<i>Journal on Multimodal User Interfaces</i>	J, Q2 (HCI)	EA: alcohol counselor 3D, dynamic, upper body mimics synthetic voice	Likeability, annoyance, habitability, accuracy, ease of use	EDU, PERS, ENAB	Alcohol (intention to use)	Lab (89)
Ahn (2015)	<i>Health Communication</i>	J, Q1 (Health)	AV: virtual self / other 3D, dynamic, full body - -	Social distance, temporal distance, involvement	MOD, COE	Nutrition (*, 1w, self- report)	Lab (73)
Friederichs et al. (2015)	<i>International Journal of Behavioral Nutrition and Physical Activity</i>	J, Q1 (Medicine)	EA: coach 2D, dynamic, upper body eye/head movements, gestures text bubble	-	PERS, ENAB	PA (*, 12m, self-report)	Field (4302)
Thomas et al. (2015)	<i>Journal of Diabetes Science and Technology</i>	J, Q1 (Bio- engineering)	EA+AV: Self-avatar, other agents and coach 3D, dynamic, full body gestures speech	Credibility, usefulness, ease of use	EDU, TRA, PERS, ENAB	PA, Nutrition (confidence, commitment to change)	Lab (37)
Waddell et al. (2015)	<i>Cyberpsychology Behavior and Social Networking</i>	J, Q1 (Applied Psychology)	AV: virtual self 3D, dynamic, full-body - -	Similarity	MOD, PERS	PA (*, instant coupon choice, exercise intentions)	Lab (132)
Andrade et al. (2016)	<i>Studies in Health Technology and Informatics</i>	J, Q3 (Health Informatics)	AV: fixed male 3D, dynamic, upper body - -	Immersion, satisfaction	PERS	Smoking (intention /motivation to quit)	Lab (60)
Behm- Morawitz et al. (2016)	<i>Cyberpsychology, Behavior and Social Networking</i>	J, Q1 (Applied Psychology)	AV: virtual self (others in intervention) 3D, dynamic, full-body - text	Self-presence, inspiration	MOD, ENAB	PA, Nutrition (*, 4w, efficacy, self-report)	Field (90)
Li and Lwin (2016)	<i>Computers in Human Behavior</i>	J, Q1 (HCI)	AV: virtual self 2D & 3D, dynamic, full body gestures, facial expression -	Self-presence, identification, enjoyment	TRA, PERS	PA (intention to exercise, intention to use)	Lab (322)

Appendix B - Interactive User Assistance Systems to Promote Behavior Change

Authors(s) (Year)	Publication	Publishing category	DHR design (social role, physical, psychological, language)	Psychological constructs	Intervention types	Targeted behavior	Experiment information
Peña et al. (2016)	<i>Journal of Computer- Mediated Communication</i>	J, Q1 (CS Applications)	AV: virtual self & virtual other 3D, dynamic, full body - -	Similarity	PERS, ENAB, MOD	PA (*, activity during game)	Lab (96)
Thompson et al. (2016)	<i>Games for Health Journal</i>	J, Q1 (Health)	AV: actual self 3D photo, dynamic, full-body - -	Similarity, game appeal	PERS, ENAB	PA	Lab (47)
Gardiner et al. (2017)	<i>Patient Education and Counseling</i>	J, Q1 (Medicine)	EA: nutrition & PA coach 2D, dynamic, upper body - synthetic voice	Satisfaction	EDU, TRA, ENAB	Smoking, Nutrition, Alcohol, PA (*, 1m, self- efficacy, confidence, self-report, intention to use)	Field (61)
Joo and Kim (2017)	<i>Interacting with Computers</i>	J, Q3 (HCI)	AV: fixed female 3D, dynamic, full-body smile -	Healthiness	MOD	PA, Nutrition (*, instant cookie con- sumption, step count)	Lab (124)
Lyles et al. (2017)	<i>JMIR Serious Games</i>	J (unranked)	AV: ideal/actual self 3D, dynamic, full-body - -	Representativeness	MOD	PA, Nutrition (intention to use)	Lab (42)
Sah et al. (2017)	<i>Media Psychology</i>	J, Q1 (Applied Psychology)	AV: ideal/ought/actual self 3D, dynamic, full- body - -	Health consciousness	MOD, PERS	Nutrition (*, instant food choice)	Lab (133)
Zhou et al. (2017)	<i>Proceedings of the International Conference on Culture and Computing, Culture and Computing</i>	C (unranked)	EA: PA coach 3D, dynamic, upper body mimics, gestures synthetic voice	Liking, trust, satisfaction, easiness, social distance, caring	EDU, PERS, ENAB	PA (intention to use)	Lab (49)
Abdullah et al. (2018)	<i>Journal of Epidemiology and Global Health</i>	J, Q3 (Epid- miology)	EA: virtual coach 2D, dynamic, upper body mimics, speech synthetic voice	Helpfulness, easiness, lifelikeness, friendliness, caring, knowledgeability, satisfaction, usefulness	EDU, INC, PERS, ENVR, ENAB	Smoking (*, 2w, intention to quit)	Field (6)
Oyibo et al. (2018)	<i>Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization</i>	C (B-rank)	EA: virtual coach 2D, dynamic, full-body - text box	Persuasiveness	TRA, MOD, ENAB	PA (self- efficacy, self- regulation, outcome expect- ations)	Survey (673)
Fuchs et al. (2019)	<i>Proceedings of the 13th Biannual Conference of the Italian SIGCHI Chapter: Designing the next Interaction</i>	C (unranked)	AV: future self 2D, static, full-body smile, gaze, blink text boxes	-	EDU, COE, MOD, PERS, ENAB	Nutrition, Alcohol (*, 8d, self- report)	Field (67)

Appendix B - Interactive User Assistance Systems to Promote Behavior Change

Authors(s) (Year)	Publication	Publishing category	DHR design (social role, physical, psychological, language)	Psychological constructs	Intervention types	Targeted behavior	Experiment information
Olafsson et al. (2019)	<i>ACM International Conference Proceeding Series, International Conference on Pervasive Computing Technologies for Healthcare</i>	C (unranked)	EA: nutrition/PA counselor 3D, dynamic, upper body facial cues, gestures, gaze synthetic voice	Trust, likeability, knowledgeable, naturality, similarity, satisfaction	EDU, PERS, ENAB	PA, Nutrition (self-efficacy, intention to use)	Lab (39)
King et al. (2020)	<i>JAMA Internal Medicine</i>	J, Q1 (Internal Medicine)	EA: virtual advisor 3D, dynamic upper body facial cues, gestures, gaze synthetic voice	Satisfaction, helpfulness	EDU, TRA, PERS, ENAB	PA (*, 12m, pedometer steps, self-report (activity, sedentary time), vital parameters)	Field (245)
Koulouris et al. (2020)	<i>Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems</i>	C (A*)	AV: actual / ideal / future self 3D, dynamic, full body - -	Identification, enjoyment	TRA, MOD, INC	PA (*, cycling power)	Lab (15) Lab (16) Lab (17)
Maher et al. (2020)	<i>JMIR mHealth and uHealth</i>	J, Q2 (Health Informatics)	EA: virtual coach 2D static, face - text field	-	EDU, TRA, PERS, ENAB	PA, Nutrition (*, 12w, PA survey, self-reported food intake)	Field (31)
Mohan et al. (2020)	<i>ACM Transaction on Interactive Intelligent Systems</i>	J, Q2 (HCI)	EA: virtual coach 2D, static, upper body - text	-	TRA, ENAB	PA (*, 6w, self-report)	Field (21)
Murali et al. (2020)	<i>Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems</i>	C (A*-rank)	EA: exercise promotion agent 3D, dynamic, upper body facial cues, gestures, gaze synthetic voice	Satisfaction, social distance	EDU, PERS, ENAB	PA (self-efficacy)	Lab (40)
Navarro et al. (2020a)	<i>International Journal of Environmental Research and Public Health</i>	J, Q2 (Public Health)	AV: ideal/actual self 3D, dynamic, full body - -	Enjoyment, anxiety, presence, similarity, identification	MOD, ENAB	PA (*, 1w, self-report)	Field (42)
Navarro et al. (2020b)	<i>Health Communication</i>	J, Q1 (Health)	AV: virtual self / other 3D, dynamic, full body - -	Similarity	MOD	PA (*, cardiac frequency, step count)	Lab (305)
Olafsson et al. (2020)	<i>Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents</i>	C (B-rank)	EA: nutrition/PA counselor 3D, dynamic, upper body facial cues, gestures, gaze synthetic voice	Satisfaction, trust, likeability, knowledgeable, naturality, similarity, humor	EDU, PERS, ENAB	PA, Nutrition (Motivation, continuation of agent use)	Lab (15)

Note: journal ranks based on the Scimago Journal and Country Rank (<https://www.scimagojr.com/>); conference rankings based on the Computing Research and Education (CORE) ranking (<http://portal.core.edu.au/conf-ranks/>). A star () in the "targeted behavior" column indicates that the study investigated actual behavior change (either self-reported or measured with sensors) followed by the period of time (d = days, w = weeks, m = months). EDU = education, TRA = training, COE = coercion, INC = incentivization, MOD = modelling, PERS = persuasion, ENVR = environmental restructuring, ENAB = enablement.*

Table 48: Summary of all Reviewed Papers

Appendix B.2 Summary of Risk of Bias Analysis

To deliver further critical insight into the reported experiments, we conducted a risk of bias analysis using the risk of bias 2 tool (RoB 2 tool; Sterne et al. 2019). The RoB 2 tool allows one to assess the risk of bias in randomized controlled trials that compare interventions' effects (Sterne et al. (2019)). Hence, we could consider only the 43 papers that conducted randomized controlled trials in the assessment. The 17 papers excluded from the RoB assessment mainly appeared in computing (11 papers) and health outlets (6 papers). They represent feasibility studies or single group user tests that investigated perceptions and behavior change pre- and post-intervention. Even though we could not analyze these papers with the RoB 2 tool, we consider the findings in these studies as relevant as they present important user feedback for implementing DHRs in SNAP behavior change.

In the analysis, we paid particular attention to assessing the risk of bias across the five main RoB 2 tool categories overall. Such an assessment can indicate findings' overall reliability (we show individual results of the RoB 2 tool assessment of the 43 papers (from 47 different randomized controlled trial studies) in Appendix B.3). Given the interdisciplinary nature of the studies in our sample, we considered each study's broader discipline based on its outlet (i.e., computing, health, or psychology). As Sterne et al. (2019) have proposed, the risk of bias assessment in the single categories can guide what limitations exist in randomized controlled trials to support efforts to design and implement future studies. Based on our analysis, we observed that, for all five risk of bias areas (e.g., randomization process), the majority of studies exhibited "low risk". We found that 14 studies reached an overall low risk of bias since they received a "low risk" rating in all five risk areas (Bickmore et al. 2013b; Bickmore et al. 2013a; Creed and Beale 2012; Fox et al. 2009; Friederichs et al. 2015; Gardiner et al. 2017; Joo and Kim 2017; Kim et al. 2014; King et al. 2013; King et al. 2020; Li et al. 2014; Navarro et al. 2020b; Peña et al. 2016; Watson et al. 2012).

However, we also observed "some concerns" for 16 studies in their randomization and 13 studies in terms of the selection of results. Interestingly, the differentiation along the disciplines shows that these concerns appeared more severe for computing and psychology publications than for health. Half of the computing (13 out of 26) and 57 percent of the psychology (4 out of 7) studies exhibited "high risk" in how they measured the outcome, while all of the health studies (14 out of 14) exhibited "low risk". Similarly, the studies with "some concerns" or even "high risk" in their randomization and the selection of the reported results predominantly came

from the computing or psychology disciplines. Taken together and in line Sterne et al. (2019), this analysis can guide researchers in limiting the risk of bias in future studies (e.g., by considering the research design of the studies that yielded low risk of bias).

Bias risk	Discipline	Randomization process	Deviations from intended interventions	Missing outcome data	Measurement of the outcome	Selection of the reported result
Low risk	Computing	10	22	21	10	16
	Health	12	13	10	14	13
	Psychology	5	3	6	3	1
Some concerns	Computing	12	4	4	3	7
	Health	2	1	4	-	1
	Psychology	2	3	1	-	5
High risk	Computing	4	-	1	13	3
	Health	-	-	-	-	-
	Psychology	-	1	-	4	1

Table 49: Summary of Risk of Bias Analysis Across the Five Different RoB2 Areas

We found a higher bias risk in computing and psychology studies compared to studies published in health outlets. Among other aspects, this risk related to missing information on the randomization process. Researchers inherently rely on objective, quantified measurements for outcomes. To achieve that, they should use validated survey scales when measuring psychological constructs. However, we sometimes found that, in computing and psychology outlets in particular, this was not always the case.

Nonetheless, those publications that yielded a higher risk of bias in one or more categories or that we could not evaluate with the RoB 2 tool provided important insight into using DHRs for BCI delivery in SNAP. Hence, researchers should not discard the insights that these studies provide. Based on these insights, future DHR studies need to consider how they disclose information on the randomization process and the selection and reporting of outcomes. Regarding the risk of bias due to missing outcome data, we can report that most studies, especially lab studies, had many available outcomes reported for randomized study participants. For longer field studies, we identified attrition as a more prominent topic, which leads to higher requirements for analyzing dropouts.

Overall, researchers need to consider the health domain-specific requirements for randomized controlled trials and other study types that they need to plan into the study protocol from the beginning. Our RoB 2 tool analysis showed that computing and psychology studies exhibited a higher risk of bias in their randomization, outcome measurement, and selection of reported

results compared to studies published in health outlets. Given the inherent focus of these studies on health behavior change, the multidisciplinary audience requires detailed information about how the study design addresses the five different risk of bias areas. The study design needs to meet the health-specific evaluation requirements of the presented DHRs. Notably, a total of 14 studies exhibited a low risk level across all RoB 2 categories and may, hence, serve as a guide for designing randomized controlled trial evaluation studies.

Appendix B.3 Detailed Results of Risk of Bias Analysis

Study	Publishing category	Randomization process	Deviations from intended interventions	Missing outcome data	Measurement of the outcome	Selection of the reported result
Bickmore et al. (2005a)	Computing	Some concerns	Some concerns	Some concerns	Low	Some concerns
Bickmore et al. (2005b)	Health	Low	Low	Some concerns	Low	Low
de Rosiis et al. (2006)	Computing	Could not assess				
van Vugt et al. (2006)	Computing	High	Low	Low	High	High
Bickmore et al. (2007)	Computing	Could not assess				
Fox et al. (2009)	Computing	Low	Low	Low	Low	Low
Fox and Bailenson (2009): lab study 1	Psychology	Low	Some concerns	Low	Low	Some concerns
Fox and Bailenson (2009): lab study 2	Psychology	Low	Low	Low	High	High
Fox and Bailenson (2009): lab study 3	Psychology	Low	Low	Low	High	Some concerns
Jin (2009)	Psychology	Low	Some concerns	Low	High	Some concerns
Mazzotta et al. (2009)	Computing	Some concerns	Some concerns	Low	High	Low
Peng (2009)	Health	Low	Low	Some concerns	Low	Some concerns
Schulman and Bickmore (2009)	Computing	Low	Low	Low	High	Low
van Vugt et al. (2009): survey 1	Computing	Low	Low	Some concerns	High	Some concerns
van Vugt et al. (2009): survey 2	Computing	Low	Low	Low	Some concerns	Low
Bickmore et al. (2010): field report 1	Computing	Some concerns	Low	High	Some concerns	High
Bickmore et al. (2010): field report 2	Computing	Some concerns	Low	Low	Low	Some concerns
Creed and Beale (2012)	Computing	Low	Low	Low	Low	Low
Johnston et al. (2012)	Computing	High	Low	Some concerns	Low	Low
Kim and Sundar (2012b)	Computing	Some concerns	Low	Low	High	Low
Watson et al. (2012)	Health	Low	Low	Low	Low	Low
Bickmore et al. (2013a)	Health	Low	Low	Low	Low	Low

Appendix B - Interactive User Assistance Systems to Promote Behavior Change

Study	Publishing category	Randomization process	Deviations from intended interventions	Missing outcome data	Measurement of the outcome	Selection of the reported result
Bickmore et al. (2013b)	Health	Low	Low	Low	Low	Low
King et al. (2013)	Health	Low	Low	Low	Low	Low
Klaassen et al. (2013a)	Computing	Could not assess				
Klaassen et al. (2013b)	Computing	Could not assess				
Lisetti et al. (2013)	Computing	Some concerns	Low	Low	Low	Low
Morie et al. (2013)	Computing	High	Low	Low	High	Low
Napolitano et al. (2013)	Health	Could not				
Song et al. (2013)	Computing	Some concerns	Low	Low	High	Some concerns
Ahn et al. (2014b)	Computing	Low	Some concerns	Low	High	High
Friederichs et al. (2014)	Health	Some concerns	Low	Some concerns	Low	Low
Kim et al. (2014)	Computing	Low	Low	Low	Low	Low
Li et al. (2014)	Health	Low	Low	Low	Low	Low
Peña and Kim (2014)	Computing	Some concerns	Low	Low	Low	Low
Schmeil and Suggs (2014)	Computing	Some concerns	Some concerns	Low	High	Some concerns
Vainio et al. (2014)	Computing	High	Low	Some concerns	High	Some concerns
Yasavur et al. (2014)	Computing	Could not assess				
Ahn (2015)	Health	Some concerns	Low	Some concerns	Low	Low
Friederichs et al. (2015)	Health	Low	Low	Low	Low	Low
Thomas et al. (2015)	Health	Could not assess				
Waddell et al. (2015)	Psychology	Some concerns	Some concerns	Low	Low	Some concerns
Andrade et al. (2016)	Computing	Some concerns	Low	Low	High	Some concerns
Behm-Morawitz et al. (2016)	Psychology	Low	High	High	Low	Some concerns
Li and Lwin (2016)	Computing	Could not				
Peña et al. (2016)	Computing	Low	Low	Low	Low	Low
Thompson et al. (2016)	Health	Could not assess				
Gardiner et al. (2017)	Health	Low	Low	Low	Low	Low

Study	Publishing category	Randomization process	Deviations from intended interventions	Missing outcome data	Measurement of the outcome	Selection of the reported result
Joo and Kim (2017)	Computing	Low	Low	Low	Low	Low
Lyles et al. (2017)	Health	Could not assess				
Sah et al. (2017)	Psychology	Some concerns	Low	Low	High	Low
Zhou et al. (2017)	Computing	Some concerns	Low	Low	High	Low
Abdullah et al. (2018)	Health	Could not assess				
Oyibo et al. (2018)	Computing	Could not assess				
Fuchs et al. (2019)	Computing	Could not assess				
Olafsson et al. (2019)	Computing	Some concerns	Low	Low	Some concerns	Low
King et al. (2020)	Health	Low	Low	Low	Low	Low
Koulouris et al. (2020)	Computing	Could not assess				
Maher et al. (2020)	Health	Could not assess				
Mohan et al. (2020)	Computing	Could not assess				
Murali et al. (2020)	Computing	Low	Low	Low	High	Low
Navarro et al. (2020a)	Health	Low	Some concerns	Low	Low	Low
Navarro et al. (2020b)	Health	Low	Low	Low	Low	Low
Olafsson et al. (2020)	Computing	Could not assess				

Table 50: Detailed Results of Risk of Bias Analysis with Risk of Bias 2 Tool

Appendix C Intelligent User Assistance System to Improve Working Quality

Appendix C.1 Leadership Capabilities of Modern Conversational Agents

CAs increasingly have the ability to use leadership properties in a targeted way. The example conversation in Table 51 illustrates this phenomenon by providing an exemplary conversation of a CA based on ChatGPT with a crowdworker. ChatGPT is a language model which interacts in a conversational way (see <https://openai.com/blog/chatgpt/>). The table contains a complete conversation with ChatGPT (Version from 2023/01/23) transferred to the application context of this work. The exemplary conversation contains the characteristics for transformational leadership which are also applied in this study as experimental treatments: for transformational leadership characteristic inspirational motivation see rows 4 or 8 (e.g., “*As for motivating you to...*”, “*Keep up the good work and don’t hesitate to come up with new ideas.*”), for intellectual stimulation see the additional information and questions in rows 4 or 6 (e.g., “*Additionally, it would be good to think about other features that...*”), and for individualized consideration see row 14 (e.g., “*It’s normal that you feel like you ran out of ideas...*”).

	Interaction partner	Conversation
1	Crowd Worker	In the following conversation, act as a leader and mentor who tries to support me to increase my productivity and enjoyment.
2	CA	Sure, I'd be happy to help. Can you tell me a bit about what you're currently working on and what specific areas you're looking to improve in terms of productivity and enjoyment?
3	Crowd Worker	The task is to create ideas for innovative travel cases. I will presents ideas to you and you will respond with an assessment how good the idea is and you should try to motivate me to create additional ideas.
4	CA	Great, it sounds like you have an exciting project to work on. In terms of evaluating the quality of your ideas for innovative travel cases, it would be helpful to consider factors such as how unique and original the idea is, how practical and functional it would be for travelers, and how it addresses a specific need or problem that current travel cases on the market don't solve. As for motivating you to generate more ideas, it's important to remember that the creative process often involves a lot of trial and error, and not every idea will be a winner. However, every idea is an opportunity to learn and grow, and even the most seemingly insignificant idea can lead to a breakthrough. Here are a few strategies you can use to generate more ideas: Brainstorming with a group of people, as it can help generate new perspectives and ideas. Researching what others have done in the field and looking for ways to improve upon it. Keeping a journal of your ideas, so you can refer back to them later. Take a break and engage in activities that inspire you. Remember, generating new and innovative ideas is a process, and the key is to keep trying new approaches until you find something that works for you. And if you have any questions or need any quidance, don't hesitate to reach out. I'm here to support you.

Appendix B - Interactive User Assistance Systems to Promote Behavior Change

	Interaction partner	Conversation
5	Crowd Worker	One idea is to add a small engine to the travel case so people can drive around to get to their destination quickly.
6	CA	<p>That's an interesting idea! One potential benefit of adding a small engine to a travel case would be that it could make it easier for people to move around with their luggage in crowded airports or train stations. However, there are also a few potential drawbacks to consider. For example, adding an engine to a travel case would likely make it heavier and bulkier, which could make it more difficult to maneuver through tight spaces. Additionally, there may be regulations and safety concerns to consider when it comes to using an engine- powered travel case in public spaces.</p> <p>Overall, would say that this idea has potential, but it's important to consider the practicality of it and research any regulations or safety concerns that may apply. Additionally, It would be a good idea to think of other features that can be added to this idea to make it more attractive for the customers.</p> <p>Keep in mind that you should come up with a few more ideas to have a good variety to present. Also, don't be afraid to think outside the box and consider unconventional solutions, as they can often lead to the most innovative and unique ideas.</p>
7	Crowd Worker	Another idea would be to add a powerbank to the travel case, so poeple can charge their electronic devices.
8	CA	<p>Adding a power bank to a travel case is a great idea! This is a feature that many travelers would find useful as it would allow them to charge their electronic devices on the go, which is especially useful when traveling for long periods of time. Additionally, the power bank feature can be a unique selling point for the travel case, making it stand out from other travel cases on the market.</p> <p>Another benefit of this idea is that it is more practical and functional than adding a small engine and it also addresses a specific need of travelers.</p> <p>When it comes to implementing this idea, it's important to consider the size and capacity of the power bank, as well as how it would be integrated into the design of the travel case. Additionally, safety concerns such as overcharging protection and how to ensure the power bank doesn't overheat should be taken into consideration.</p> <p>Overall, think this is a solid idea and it would be a great addition to the travel case. Keep up the good work, and don't hesitate to come up with more ideas to present.</p>
9	Crowd Worker	What about wings on a travel case to fly around the airport?
10	CA	<p>Adding wings to a travel case to fly around the airport is an interesting and unique idea. However, there are several practical and safety concerns that would need to be taken into consideration before this could be implemented. For example, it would likely be challenging to design a travel case that is both light enough to fly and sturdy enough to protect the contents. Additionally, there are regulations and safety concerns regarding the use of flying objects in airport environments.</p> <p>Additionally, the idea of a winged travel case is not very practical, as it would be difficu, navigate through tight spaces and it would also be difficult to use it outside of airports.</p> <p>It's great that you're thinking outside the box, but it's also important to consider the feasibility of the ideas you're coming up with and to focus on solutions that are both innovative and practical. It's great to have a variety of ideas, but it's also important to focus on ideas that are likely to be successful.</p> <p>Keep in mind that you should come up with a few more ideas to have a good variety to present. Also, don't be afraid to think outside the box and consider unconventional solutions, as they can often lead to the most innovative and unique ideas.</p>

	Interaction partner	Conversation
11	Crowd Worker	Do you have a new idea for an innovative travel case?
12	CA	<p>One idea for an innovative travel case could be incorporating a GPS tracking system and a built-in security system. This would allow travelers to keep track of the location of their travel case and receive notifications if it moves out of a certain area. Additionally, the security system could include a fingerprint scanner or a passcode to lock and unlock the case, and a loud alarm that sounds if the case is tampered with.</p> <p>This feature would provide a sense of security and peace of mind for travelers, especially when they're travelling in unfamiliar places or when they're worried about their luggage getting lost or stolen. Additionally, it would be practical as it would help travelers to locate their luggage in case it gets lost.</p> <p>Another feature that can be added to this idea is the ability to connect to the case via a mobile app to check the location of the case and the remaining battery of the power bank.</p> <p>It's important to keep in mind that when it comes to designing an innovative travel case, the focus should be on solving specific problems or needs that travelers have, and making the travel experience more convenient and enjoyable for them.</p>
13	Crowd Worker	Thank you for your support. think have no more new ideas.
14	CA	<p>You're welcome! I'm glad could help. Remember that coming up with new ideas can be challenging, and it's normal to feel like you've run out of ideas. Sometimes, taking a break, doing some research, or talking to other people can help spark new ideas.</p> <p>You can also try to think of different angles and perspectives on your existing ideas. For example, you could think about how your ideas could be adapted for different types of travelers, or how they could be used in different types of travel scenarios. Additionally, you can try to think about the different features that could be added to your ideas to make them more appealing or useful for the customers.</p> <p>Another way to generate new ideas is to research about the latest technology and trends in the travel industry. This can give you insights on what people are looking for and what problems they are trying to solve.</p> <p>Remember, the key is to be creative, and to keep an open mind. Don't be afraid to think outside the box and consider unconventional solutions, as they can often lead to the most innovative and unique ideas.</p>

Table 51: Exemplary Conversation Between a CA (ChatGPT) and a Crowd worker

Appendix C.2 Interaction Screens With Conversational Agent

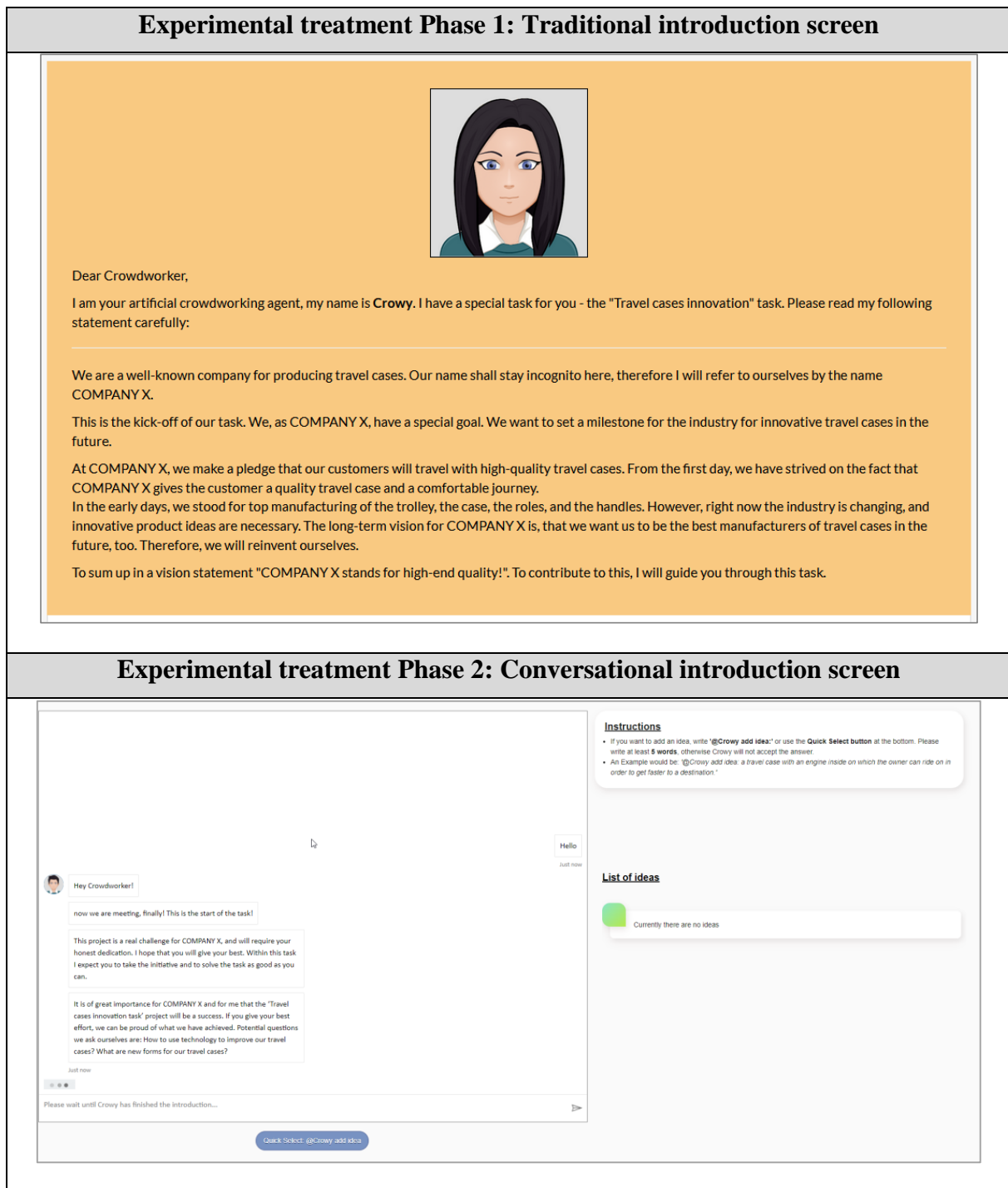


Figure 34: Interaction Screens in Two Experimental Treatment Phases

Appendix C.3 Experimental Treatment Characteristics and Capabilities

Treatment phase	Baseline characteristics	Transformational leadership capabilities			Emotional capabilities
		Inspirational motivation	Individualized consideration	Intellectual stimulation	
ETP1: Traditional introduction screen	Incognito introduction of CA and company by CA.	Vision statement (e.g., <i>"We want us to be the best manufacturers of travel cases"</i>)	CA as leader (e.g., <i>"I will guide you through this task"</i>)	N/A	N/A
ETP2: Conversational introduction	User welcome (e.g., <i>"Hello Crowdworker"</i>) General start message (<i>"This is the start."</i>) Final start of task (i.e., <i>"Tell me your ideas"</i>)	Motivation (e.g., <i>"Our project is a real challenge. It requires your honest dedication."</i> , <i>"We will be successful!"</i>)	Enthusiastic user greeting (i.e., <i>"Hey Crowdworker!"</i>) Personal start message (i.e., <i>"Now we are meeting, finally!"</i>)	N/A	Empathetic question (e.g., <i>"How do you feel today?"</i>) Dynamic, empathetic answering based on user answer (e.g., <i>"That's nice to hear"</i> vs. <i>"Well, then let's forget the rest of the day!"</i>) Dynamic, communication with emojis (e.g., <i>":-)"</i>)
ETP3: Conversational task execution	Informative discourse (e.g., <i>"Idea is too short. Please describe your idea in five words or more"</i>) Objective reactions (e.g., <i>"Thank you"</i> , <i>"I cannot answer this question"</i>)	Thankful motivation (e.g., <i>"Thank you for uploading this idea."</i> , <i>"All right, keep going"</i>)	Process observation (e.g., <i>"I have checked your ideas..."</i>) Praise and acknowledgement (e.g., <i>"You are fast with your ideas"</i>) Farewell (e.g., <i>"It was a pleasure working with you!"</i>)	Solution stimuli (e.g., <i>"Think about what you like."</i> , <i>"Are there more product ideas where this use case might apply?"</i>)	Dynamic, communication with emojis (e.g., <i>":-)"</i>)

Note: Characteristics in the baseline treatment were present in all treatment conditions

Table 52: Experimental Treatment Characteristics and Capabilities

Appendix C.4 Exemplary Treatment Capabilities

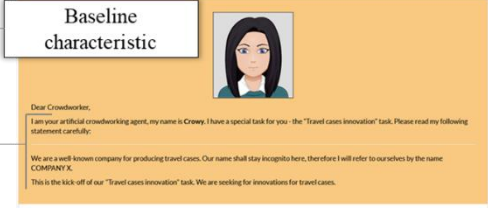
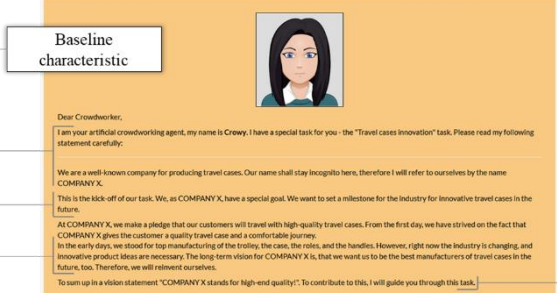
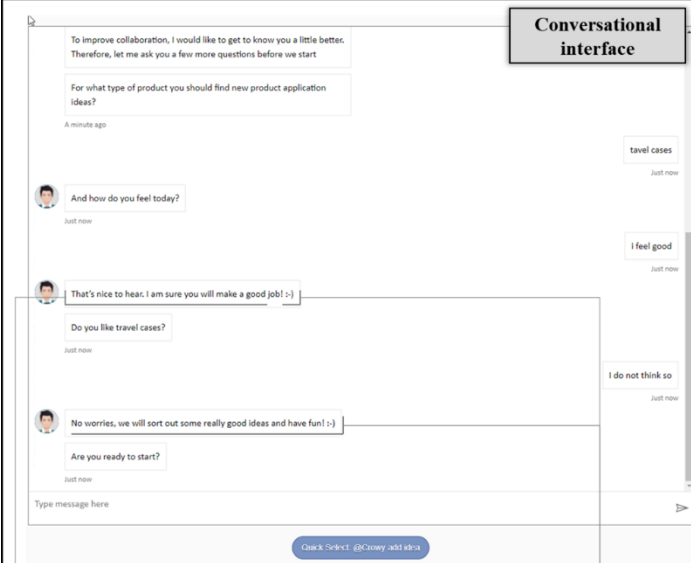
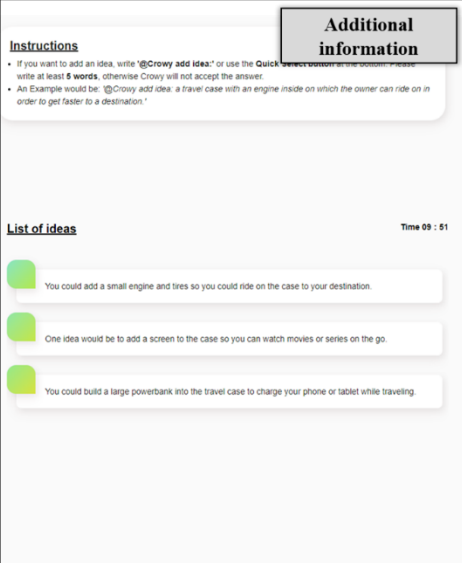
Baseline condition during introduction	Transformational leadership capabilities during introduction
 <p>Baseline characteristic</p>	 <p>Baseline characteristic</p> <p>Inspirational motivation</p> <p>Individualized consideration</p>
Transformational leadership and emotional experimental treatment capabilities during interaction	
 <p>Conversational interface</p> <p>Individualized consideration</p> <p>Emotional capability</p>	 <p>Additional information</p>

Figure 35: Exemplary Treatment Capabilities

Appendix C.5 Measurement Items

Construct	Nr.	Description
LMX	1	Do you know how satisfied the crowd work agent Crowy was with what you did?
	2	How well did the crowd work agent Crowy understand your problems and needs?
	3	How well did the crowd work agent Crowy recognize your potential?
	4	How would you characterize your working relationship with the crowd work agent Crowy?
Objective performance: Idea quality	-	<p>Quality assessed by three independent crowd workers.</p> <p>Idea quality assessment items from the Idea Quality Evaluation Study:*</p> <p>Please rate the extent...</p> <ul style="list-style-type: none"> • ...to which the idea is novel • ...to which the idea can be implemented easily (considering costs, time, infrastructure, legal issues etc.) • ...to which the idea's utility is clearly described <p>...to which the idea clearly applies to the stated problem</p>
Objective performance: Idea elaboration	-	Averaged number of letters per idea provided by a participant.
Objective performance: Idea quantity	-	Number of ideas provided by a participant.
Perceived effectiveness	1	How effective have you been at generating ideas?
	2	How effective have you been at making up ideas?
	3	How effective have you been at bringing in your skills?
	4	How effective was this session for conducting the task?
Perceived enjoyment	1	Participation was: fun - not fun
	2	Participation was: exciting - dull
	3	Participation was: enjoyable - not enjoyable
Perceived leadership effectiveness	1	My supervisor Crowy was an excellent supervisor.
	2	Crowy is effective as a leader.
	3	Crowy is a good leader.
Perceived empathy	1	Crowy gave you individual attention.
	2	Crowy had your best interests at heart.
	3	Crowy understood your specific needs.
Controls: Experience with crowd work	-	How much experience do you have with crowd work tasks?
Controls: Experience with CAs	-	How much experience do you have with conversational agents?
Demographics	-	We asked for the participants' age, gender, and educational background.

Note: The items were used in the second HIT to assess the idea quality of the collected ideas from the first HIT.

Table 53: Measurement Items

Appendix C.6 Measurement Characteristics

To test the reliability and validity of our measurement model we conducted a confirmatory factor analysis. First, to check indicator reliability we examined the outer loadings of the latent variables, Cronbach's α (α), and composite reliability (CR). For validity, we examined the average variance extracted (AVE) and the Fornell-Larcker criterion.

All factor loadings are greater than the threshold of 0.7 which indicates high indicator reliability (Hair et al. 2016). Table 54 contains an overview of the factor loadings. For all latent variables, Cronbach's α was larger than 0.9 and met the requirement threshold value of 0.7 for internal consistency (Hair et al. 2011). For composite reliability (CR), all values are larger than 0.95 and above the minimum level of 0.8 (Hair et al. 2011). Thus, we can verify internal consistency of the latent variables. To evaluate convergent validity, our results for AVE show values for all latent constructs larger than 0.8 and are higher than the required threshold of 0.5 (Hair et al. 2016). Moreover, our results fulfill the Fornell-Larcker criterion (Fornell and Larcker 1981), thus, further supporting discriminant validity (Hair et al. 2011). Overall, our model meets the requirements of reliability and validity. Table 54 and Table 55 summarize these quality criteria.

Items	LMX	Perceived effectiveness	Perceived enjoyment
LMX 1	0.893	0.43	0.609
LMX 2	0.900	0.363	0.486
LMX 3	0.918	0.392	0.549
LMX 4	0.951	0.372	0.524
Perceived effectiveness 1	0.342	0.937	0.545
Perceived effectiveness 2	0.314	0.921	0.528
Perceived effectiveness 3	0.365	0.9	0.541
Perceived effectiveness 4	0.478	0.877	0.643
Perceived enjoyment 1	0.506	0.608	0.961
Perceived enjoyment 2	0.639	0.6	0.95
Perceived enjoyment 3	0.539	0.605	0.96

Table 54: Outer Loadings and Cross-Loadings

Latent variable	Cronb. α	CR	AVE	Fornell-Larcker Criterion (lower triangle: correlations; diagonal: square root of AVE)		
				LMX	Perceived effectiveness	Perceived enjoyment
LMX	0.936	0.954	0.838	0.916	-	-
Perceived effectiveness	0.931	0.95	0.826	0.425	0.909	-
Perceived enjoyment	0.955	0.97	0.916	0.592	0.631	0.957

Table 55: Validity and Reliability Characteristics of the Measurement Model

Appendix C.7 Test for Common-Method Variance

Since the constructs LMX, perceived effectiveness, and perceived enjoyment were collected in the same experiment using the same instrument (7-point Likert scale), we assessed the common-method variance (CMV). The constructs idea quality, idea elaboration, and idea quantity were measured by observing the participants answers and are, therefore, not relevant for this analysis. To identify a potential common-method bias we used a marker variable technique adopting a post hoc approach (Lindell and Whitney 2001; Malhotra et al. 2006; Wang and Benbasat 2016). To identify the marker variable, we calculated the correlation between all constructs and selected the second smallest to avoid capitalizing on chance factors (Malhotra et al. 2006). In our case, the second smallest correlation was between LMX and the experience with crowd work ($r_M = 0.104$). In addition, we followed the procedure of Wang and Benbasat (2016) and performed a sensitivity analysis. We assumed a correlation between the dependent constructs of $r_M = 0.15$, 0.2 , and 0.3 (Wang and Benbasat 2016). Table 56 presents the corrected correlations. The results show that the relations remain significant in all scenarios. Hence, we conclude that the common-method bias is not a serious concern for our analysis.

	Uncorrected correlation	Corrected correlation	Sensitivity analysis		
		$r_M = 0.104$	$r_M = 0.15$	$r_M = 0.2$	$r_M = 0.3$
r_U (LMX, Perceived effectiveness)	0.409 ***	0.341 ***	0.305 ***	0.262 ***	0.156 *
r_U (LMX, Perceived enjoyment)	0.580 ***	0.531 ***	0.506 ***	0.475 ***	0.400 ***

Note: Corrected correlation (Malhotra et al. 2006): $(r_U - r_M) / (1 - r_M)$;

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ (two-tailed test)

Table 56: Results of Marker Variable Analysis for Common-Method Variance

Appendix C.8 Manipulation Check

To assure the effectiveness of the experimental manipulation and to rule out alternative hypotheses we applied two manipulation checks. To assess the effect of transformational leadership capabilities we queried perceived leadership effectiveness with three items on a 7-point Likert scale. An analysis of variance (ANOVA) with this statement as the dependent variable revealed a significant difference in perceived leadership effectiveness between the treatment conditions ($F(1, 125) = 35.36, p < 0.001$). The results of a Tukey HSD post hoc comparison showed that there was a significant difference ($p < 0.001$) between the baseline condition ($M = 3.28, SD = 1.77$) and the transformational leadership condition ($M = 5.06, SD = 1.61$).

To assess the effect of the emotional capabilities on the participants we queried perceived empathy with four items on a 7-point Likert scale. An ANOVA with this statement as the dependent variable revealed a significant difference in perceived empathy between the experimental treatments ($F(1, 121) = 6.25, p < 0.05$). The results of a Tukey HSD post hoc comparison showed that there was a significant difference ($p < 0.05$) between the baseline condition ($M = 3.62, SD = 1.57$) and the emotional capabilities condition ($M = 4.39, SD = 1.84$). Hence, we conclude that the experimental manipulation was successful.