

Bots at the Frontlines: Insights on How Conversational Interfaces Shape User Experience and Service Outcomes

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

submitted by

Naim Zierau

from

Germany

Approved on the application of

Prof. Dr. Jan Marco Leimeister

and

Prof. Dr. Matthias Söllner

Dissertation no. 5314

Difo-Druck GmbH, Untersiemau 2023

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

St. Gallen, October 24, 2022

The President:

Prof. Dr. Bernhard Ehrenzeller

To my parents
Schiwa and Martin

and my siblings
Shirin and Vahid

and my sister-in-law
Ricarda

and my nephews
Robin and Milan

for their love and support.

Acknowledgements

I could not have completed this dissertation without the support of many people. Over the last four years, I was very fortunate to meet a lot of bright, supportive, and inspiring people that guided and accompanied me through my dissertation journey.

First and foremost, I need to thank my advisor **Jan Marco Leimeister**. Thank you for all the scientific (and non-scientific) discussions, for encouraging me always to push forward, to explore my field, and for providing me with endless opportunities to develop as a researcher and a person.

I also need to thank my co-advisor **Matthias Söllner**, for always having an open door for me. Thank you for all the valuable and hands-on ideas and feedback you provided me over the last four years.

In addition, this dissertation would have been unimaginable without the support of my dear co-authors: **Abraham Bernstein, Anouk Bergner, Francesc Busquet, Edona Elshan, Christian Engel, Christian Hildebrand, Andreas Janson, Anuschka Schmitt, Melanie Schwede, Maik Hammerschmidt, Suzanne Tolmeijer, Camillo Visini, and Jalil Wahdatehagh**. Thank you for contributing with all the ideas, methodologies, data collections, data analyses, experimental tools, and writing that made this dissertation possible.

I also need to thank all the past and present members of the IWI and ITeG: **Torben Barev, Dennis Benner, Matthias Billert, Ivo Blohm, Ulrich Bretschneider, Ernestine Dickhaut, Philipp Ebel, Karen Eilers, Tobias Fahse, Michael Greineder, Nicolas Knop, Robin Knote, Niklas Leicht, Stefan Kleinschmidt, Sarah Oeste-Reiß, Christoph Peters, Mahei Li, Marcel Rhyn, Eva Ritz, Roman Rietsche, Thiemo Wambsgaß, Rainer Winkler, Sofia Schöbel, Benedikt Simmert, Fabian Tingelhoff, Julia Troll, Rainer Winkler, Sascha Weigel, and Tim Weinert**. Thank you for the countless hours spent together and the mutual support that made this time truly special.

I also need to thank **Susanne Gmünder**, who has always encouraged me and supported me in all sorts of tasks during the last four years (and provided me with an endless stock of chocolate).

I would also like to thank my many wonderful friends who have accompanied me throughout this time and have always supported and advised me. Thank you, **Jonathan Broekhuizen, Jan-Hendrik Bucher, Henrietta Dorf Müller, Paul Dunker, Lukas**

Falke, Bettina Hartmann, Felix Hermes, Hannah Leimert, Bayan Morocho, Boryana Milova, Lena-Marie Pätzmann, Lena Weiss, and Sophie Stopp.

I also thank both the **Swiss National Science Foundation** and the **Basis Research Fund** of the University of St.Gallen for supporting this work. This PhD dissertation has been generously supported by the **Swiss National Science Fund**, with funding obtained by Jan Marco Leimeister, on the research project “Understanding and Designing User Trust in Smart Personal Assistants” (192718), and by the **Basic Research Fund** of the University of St.Gallen, with funding obtained by Christian Hildebrand and Jan Marco Leimeister, on the research project “Voice-Controlled Interfaces as Scalable Trust Detection and Service Excellence Systems” (2200687).

I also need to thank the scientific community and the countless anonymous reviewers that helped shape the papers on which this dissertation is based.

Finally, I need to thank my whole family, who have always been there for me. They were the mental and emotional support I needed throughout this time. Thank you, **Mama & Papa, Shirin, Ricarda, Vahid, Oma Gila, Grossmutti, Astrid, Kimia, and Sahba!**

St. Gallen, June 2022

Naim Zierau

List of Dissertation Papers

PAPER A: **Zierau, N.**; Elshan, E.; Visini, C. & Janson, A. **(2020)**: A Review of the Empirical Literature on Conversational Agents and Future Research Directions. *In: International Conference on Information Systems (ICIS)*.

PAPER B: **Zierau, N.**; Engel, C.; Söllner, M. & Leimeister, J.M. **(2020)**: Trust in Smart Personal Assistants: A Systematic Literature Review and Development of a Research Agenda. *In: 15th International Conference on Wirtschaftsinformatik (WI 2020)*.

PAPER C: **Zierau, N** **(2022)**: Gendered Bot as Surrogates of Trust? How Voice Gender and Conversational Style Affect Interface Trust and Firm Perceptions. *Submitted to: Proceedings of the 56th Hawaii International Conference on System Sciences*.

PAPER D: Tolmeijer, S.; **Zierau, N.**; Janson, A.; Wahdatehagh, J. S.; Leimeister; J. M. & Bernstein, A. **(2021)**: Female by Default? – Exploring the Effect of Voice Assistant Gender and Pitch on Trait and Trust Attribution. *In: ACM CHI 2021 Late Breaking Work*.

PAPER E: **Zierau, N.**, Hildebrand, C., Bergner, A., Busquet, F., Schmitt, A. & Leimeister J.M. **(2022)**: Voice Bots at the Frontlines: Voice-Based Interface Enhance Interface Flow & Boost Service Outcomes. *Accepted for publication: Journal of the Academy of Marketing Science*

PAPER F: Schwende, M., **Zierau, N.**, Janson, A, Hammerschmidt, M. & Leimeister, J.M. **(2022)**: “I Will Follow You!” – How Recommendation Modality Impacts Processing Fluency and Purchase Intention. *Submitted to: International Conference on Information Systems (ICIS)*.

Summary of Contents

Acknowledgements	i
List of Dissertation Papers.....	iii
Summary of Contents.....	v
Table of Contents.....	vii
List of Abbreviations	xv
List of Figures	xvii
List of Tables	xix
Abstract	xxi
Kurzfassung	xxiii
Part A: Synopsis	1
1 Introduction	1
2 Theoretical Background and Related Work.....	9
3 Research Strategy	23
4 Results.....	28
5 Discussion and Outlook.....	43
Part B: Papers of the Dissertation	63
Paper A – A Review of the Empirical Literature on Conversational Agents and Future Research Directions.....	65
Paper B – Trust in Smart Personal Assistants: A Systematic Literature Review and Development of a Research Agenda	91
Paper C – Gendered Bots as Surrogates of Trust? How Voice Gender and Communication Style Affect Interface Trust and Firm Perceptions	107
Paper D – Female by Default? – Exploring the Effect of Voice Assistant Gender and Pitch on Trait and Trust Attribution	125
Paper E – Voice Bots on the Frontline: Voice-Based Interfaces Enhance Flow- Like Consumer Experiences & Boost Service Outcomes	139

Paper F – “I Will Follow You!” – How Recommendation Modality Impacts Processing Fluency and Purchase Intention	187
References	209
Curriculum Vitae	249

Table of Contents

Acknowledgements	i
List of Dissertation Papers.....	iii
Summary of Contents.....	v
Table of Contents.....	vii
List of Abbreviations	xv
List of Figures	xvii
List of Tables	xix
Abstract	xxi
Kurzfassung	xxiii
Part A: Synopsis	1
1 Introduction	1
1.1 Problem Statement	1
1.2 Research Questions and Research Methods.....	3
1.3 Structure of this Dissertation.....	7
2 Theoretical Background and Related Work.....	9
2.1 The Past and Present of Conversational Interfaces	9
2.2 Distinguishing Conversational Interfaces From Other Self-Service Channels	12
2.3 The Unique Features of Conversation-Based Service Delivery	13
2.3.1 Feature 1: Voice Gender.....	14
2.3.2 Feature 2: Interaction Modality	15
2.4 Communication Theories	16
2.4.1 Social Response Theory	17
2.4.2 Media Richness Theory.....	17
2.5 User Experience	18
2.5.1 Trustworthy User Experiences	19

2.5.2	Absorbing User Experiences.....	19
2.5.3	Fluent User Experiences	20
2.6	Conversational Design	20
2.7	Service Outcomes.....	21
2.8	Conceptual Framework	21
3	Research Strategy.....	23
3.1	Epistemological Foundation.....	23
3.2	Research Methods	24
3.2.1	Systematic Literature Review	24
3.2.2	Experimental Research.....	25
4	Results	28
4.1	Overview of Publications.....	28
4.2	Papers of the Dissertation.....	30
4.2.1	Paper A: A Review of the Empirical Literature on Conversational Agents and Future Research Directions.....	31
4.2.2	Paper B: Trust in Smart Personal Assistants – A Systematic Literature Review and Development of a Research Agenda	33
4.2.3	Paper C: Gendered Bots as Surrogates of Trust? How Voice Gender and Conversational Style Affect Interface Trust and Firm Perceptions	35
4.2.4	Paper D: Female by Default? – Exploring the Effect of Voice Assistant Gender and Pitch on Trait and Trust Attribution	37
4.2.5	Paper E: Voice Bots at the Frontlines – Voice-Based Interface Enhance Interface Flow & Boost Service Outcomes.....	39
4.2.6	Paper F: “I Will Follow You!” – How Recommendation Modality Impacts Processing Fluency and Purchase Intention	41
5	Discussion and Outlook	43
5.1	Summary of Results	43
5.2	Theoretical Implications.....	46

5.2.1	RQ1: Toward a Holistic Conceptualization of Conversational Interfaces in Service Contexts	46
5.2.2	RQ2: The Effects of Voice Gender on User Experiences and Service Outcomes	47
5.2.3	RQ3: The Effects of Interaction Modality on User Experiences and Service Outcomes	48
5.3	Managerial Implications	49
5.3.1	Toward an Effective Deployment of Conversational Interfaces	49
5.3.2	Toward an Effective Design of Voice	50
5.3.3	Toward an Effective Design of Conversational Interactions	51
5.4	Limitations	53
5.5	Future Work	54
5.5.1	Avenue 1: Conversational Interactions	56
5.5.2	Avenue 2: Designing Conversational Service Interactions	57
5.5.3	Avenue 3: Potential of Voice-Based Automated Service Channels	59
5.5.4	Avenue 4: Ethical Implications of Gendered Interfaces	60
Part B: Papers of the Dissertation		63
Paper A – A Review of the Empirical Literature on Conversational Agents and Future Research Directions		65
A.1	Introduction	66
A.2	Conceptual Background	67
A.3	Research Approach	69
A.3.1	Paper Selection Process	69
A.3.2	Paper Analysis	71
A.4	Results	72
A.4.1	Section 1: Descriptive Findings	73
A.4.2	Section 2: Findings on Dependent Variables	73
A.4.3	Section 3: Findings on Independent Variables	75

A.4.4	Section 4: Findings on the Relationship between Independent and Dependent Variables	76
A.5	Discussion and Development of a Research Agenda.....	84
A.6	Limitations	88
A.7	Conclusions	88
Paper B – Trust in Smart Personal Assistants: A Systematic Literature Review and Development of a Research Agenda		91
B.1	Introduction	92
B.2	Theoretical Foundation	93
B.2.1	Smart Personal Assistants	93
B.2.2	Trust in Information Systems.....	94
B.3	Method	95
B.4	Results	98
B.4.1	User Interface-Driven Trust in Smart Personal Assistants	99
B.4.2	Interaction-Driven Trust in Smart Personal Assistants.....	101
B.4.3	Explanation-Driven Trust in Smart Personal Assistants.....	102
B.5	Discussion and Development of a Research Agenda.....	103
B.6	Limitations	105
B.7	Conclusion.....	106
Paper C – Gendered Bots as Surrogates of Trust? How Voice Gender and Communication Style Affect Interface Trust and Firm Perceptions		107
C.1	Introduction	108
C.1.1	Trust in Automated Interface	110
C.1.2	Voice User Interfaces and Trust.....	110
C.1.3	Related Work on Trust in Voice Gender	111
C.2	Hypothesis Development	113
C.3	Research Method.....	116
C.3.1	Participants	116

C.3.2	Experimental Context.....	116
C.3.3	Experimental Procedure	116
C.3.4	Experimental Stimuli.....	117
C.3.5	Measurements.....	118
C.4	Results	119
C.5	Discussion and Contribution	121
C.6	Limitations and Future Research.....	123
Paper D – Female by Default? – Exploring the Effect of Voice Assistant Gender and Pitch on Trait and Trust Attribution		125
D.1	Introduction	126
D.2	Related Work.....	127
D.3	Method.....	129
D.3.1	Experimental Platform and Voice Design.....	129
D.3.2	Experiment Procedure	131
D.3.3	Participants	132
D.3.4	Measurement and Analysis	132
D.4	Results	133
D.4.1	Trait Ascription	133
D.4.2	Trust.....	135
D.4.3	The Role of Task Context	136
D.5	Discussion	136
D.6	Contribution and Future Work	138
D.7	Acknowledgments	138
Paper E – Voice Bots on the Frontline: Voice-Based Interfaces Enhance Flow-Like Consumer Experiences & Boost Service Outcomes		139
E.1	Introduction	140
E.2	Theoretical Background	141
E.2.1	Related Prior Work on Voice-Based Interfaces	141

E.2.2	Key Conceptual Properties of Voice-Mediated Communication	144
E.2.3	Voice-Based Interfaces and Flow	145
E.2.4	The Impact of Interface Flow on Service Experience & Firm Outcomes.....	146
E.2.5	Boundary Conditions: The Role of Semantic Fluency & Conversational Turns	147
E.3	Experimental Paradigm and Context	150
E.4	Overview of Studies	151
E.5	Study 1.....	151
E.5.1	Design and Procedure	151
E.5.2	Results	152
E.5.3	Discussion	154
E.6	Study 2.....	154
E.6.1	Design and Procedure	154
E.6.2	Results	155
E.6.3	Discussion	157
E.7	Study 3.....	157
E.7.1	Design and Procedure	158
E.7.2	Results	158
E.7.3	Discussion	160
E.8	Study 4.....	160
E.8.1	Design and Procedure	161
E.8.2	Results	162
E.8.3	Discussion	163
E.9	General Discussion.....	163
E.9.1	Theoretical implications.....	166
E.9.2	Managerial implications.....	167
E.9.3	Future Research.....	169

E.10	Web Appendix.....	170
E.10.1	A. Supplementary Methods.....	170
E.10.2	B. Supplementary Analyses	184
Paper F – “I Will Follow You!” – How Recommendation Modality Impacts Processing Fluency and Purchase Intention		187
F.1	Introduction	188
F.2	Theoretical Background and Related Work	189
F.2.1	The Role of Processing Fluency in Purchase Contexts.....	190
F.2.2	Related Work on Recommendation Modalities	191
F.2.3	Related Work on Explanations of Recommendations	192
F.3	Hypotheses Development.....	193
F.3.1	Research Framework.....	194
F.3.2	The Effect of Recommendation Modality on Perceptions and Purchase Intention	194
F.3.3	The Effect of Explanation of Recommendation on Processing Fluency	196
F.3.4	The Interplay of Recommendation Modality and Explanations	196
F.4	Experimental Procedures and Research Context	197
F.5	Study 1	198
F.5.1	Design and Procedures	198
F.5.2	Results	200
F.5.3	Discussion	200
F.6	Study 2.....	200
F.6.1	Design and Procedures	201
F.6.2	Results	201
F.6.3	Discussion	203
F.7	General Discussion.....	203
F.8	Implications, Limitations and Future Research.....	205

References 209

Curriculum Vitae 249

List of Abbreviations

AI	artificial intelligence
ANCOVA	analysis of covariance
ANOVA	analysis of variance
AVE	average variance extracted
CA	conversationa agent
CASA	computers are social actions
CI	conversational interface
CR	composite realibility
et al.	et alii, and others
etc.	et cetera, and similar other things
H	hypothesis
HCI	human-computer interaction
HICCS	Hawaii International Conference on System Sciences
i.e.	id est, in other words
ICIS	International Conference on Information Systems
IS	information systems
M	mean
NLP	natural language processing
RQ	research question
SLR	sythematic literature review
SD.	standard deviation
SE	standard error
SPA	smart personal assistant
VUI	voice user interface

List of Figures

Figure 1. Overview of research questions	3
Figure 2. Structure of this dissertation.....	8
Figure 3. Conceptual framework of this dissertation	22
Figure 4. Text-based and voice-based version of experimental platform	27
Figure 5. Overview of how papers contribute to the conceptual framework of this dissertation	30
Figure 6. Number of publications over time (n=74, included in analysis of findings).....	73
Figure 7. CA outcomes regarding rapport.....	78
Figure 8. CA outcomes regarding social presence	79
Figure 9. CA outcomes regarding trust	81
Figure 10. CA outcomes regarding utility	82
Figure 11. Number of publications over time	98
Figure 12. Research Framework.....	114
Figure 13. Experimental Platform	117
Figure 14. Interaction of Voice Gender and Conversational Style	120
Figure 15. Online voice assistant interface.....	129
Figure 16. Example excerpt of the compliance task.....	132
Figure 17. Conceptual model.....	150
Figure 18. Voice-based interfaces promote flow-like consumer experiences.....	153
Figure 19. Path model results (Study 2)	157
Figure 20. Path model results (Study 3)	160
Figure 21. Research model	194
Figure 22. Text-based vs. speech-based CA	198
Figure 23. Predictive margins of the interaction effects for processing fluency.....	202
Figure 24. Path model of Study 1 & 2.....	203

List of Tables

Table 1. Differences among prevalent service encounters (adapted from Komiak & Benbasat, 2003)	12
Table 2. Difference between website- and conversational interface-mediated service encounters (adapted from Hildebrand & Bergner, 2020).	14
Table 3. Research design decision (adapted from Recker, 2013 and Cooper & Emory, 1991).	27
Table 4. Publications included in this dissertation	28
Table 5. Summary of results across all papers.	45
Table 6. Agenda for future research on conversational interfaces in service contexts	55
Table 7. Bibliographical metadata Paper A	65
Table 8. Overview of searched journals and conference proceedings	71
Table 9. Summary of research process	84
Table 10. Bibliographical metadata Paper B	91
Table 11. Results of the literature search process.....	97
Table 12. Definition and conceptual boundaries of the three perspectives on trust in SPAs	98
Table 13. Preliminary research agenda on user trust in SPAs	104
Table 14. Bibliographical metadata Paper C	107
Table 15. Review of Relevant Literature on Voice Gender	112
Table 16. Measurements	119
Table 17. Bibliographical metadata Paper D	125
Table 18. Overview of selected voices	131
Table 19. Selected average trait ascription scores per voice pitch	135
Table 20. Bibliographical metadata Paper E	139
Table 21. Review of relevant literature on voice-based interface modalities.....	142

Table 22. Key conceptual properties of text-mediated vs. voice-mediated communication 144

Table 23. Summary of studies 165

Table 24. Bibliographical metadata Paper F 187

Abstract

Conversational Interfaces (CIs) transform how users interact with firms across a wide range of service interactions. Based on continuous advances in artificial intelligence (AI) and natural language processing (NLP), CIs allow firms to automate customer engagement with users still experiencing the dynamic of human interaction, almost as if they were talking to a service employee. From assisting users in filing insurance claims to providing product recommendations, CIs increasingly support users in solving service requests more intuitive and natural than traditional service technologies (e.g., websites). Following a novel interaction paradigm, CIs mimic a human conversation style in either written or spoken form. All this allows CIs to interact with users in what was previously considered a hallmark of human communication: natural language-based dialogue. This raises fundamental questions about how users respond to CI-facilitated service interactions and the downstream consequences on service outcomes.

This dissertation's overarching contribution is a novel framework of CI-facilitated service interactions, building on and integrating fundamental theories of human communication, technology-mediated service delivery, and the emerging work on human-AI relationships. The framework is developed and tested in six complementary paper projects. In combination, these papers demonstrate how CIs and their design (i.e., voice gender, interaction modality) fundamentally affect user experiences (i.e., trust, flow, fluency) and, ultimately, related service outcomes. Further, they assess how moderating factors manifested in the conversational design of CIs influence these effects. Considering these contributions, this dissertation has important implications for both research and practice: From a theoretical perspective, this dissertation shows how fundamental insights from human communication can be harnessed to advance the emerging and interdisciplinary research on how users relate to and are influenced by CIs in service interactions. The findings of this dissertation, likewise, have important practical implications for firms that aim to leverage the potential of CIs to improve user experiences and the theory-driven conversational design of CIs.

Keywords

Conversational Interface, Speech Interaction, Voice Assistant, User Experience, Conversational Design, Service Outcomes, Service Research

Kurzfassung

Sprachbasierte Interfaces (Conversational Interfaces, CIs) verändern die Art und Weise, wie Nutzer mit Unternehmen in einem breiten Spektrum von Dienstleistungen interagieren. CIs ermöglichen den Unternehmen, auf der Grundlage der derzeitigen kontinuierlichen Fortschritte in der künstlichen Intelligenz (KI) und der Verarbeitung natürlicher Sprache (NLP), die Kundeninteraktion zu automatisieren, ohne dass die Dynamik menschlicher Interaktion verloren geht. Somit erzeugen CIs eine Dienstleistungserfahrung, die der Interaktion mit einem menschlichen Servicemitarbeiter ähnelt. Von der Unterstützung bei der Anmeldung von Versicherungsansprüchen bis hin zur Bereitstellung von Produktempfehlungen unterstützen CIs die Nutzer zunehmend bei der Lösung von Serviceanfragen. Sie sind dabei intuitiver und natürlicher als herkömmliche Servicetechnologien (z. B. Websites). Einem neuartigen Interaktionsparadigma folgend ahmen CIs zwischenmenschliche Kommunikation in schriftlicher oder gesprochener Form nach. Als Folge können CIs mit den Nutzern auf eine Art und Weise interagieren, die bisher als Wesensmerkmal menschlicher Kommunikation galt: ein auf natürlicher Sprache basierender Dialog. Dies wirft grundlegende Fragen darüber auf, wie Nutzer auf CI-gestützte Service-Interaktionen reagieren und welche Auswirkungen dies auf das Serviceergebnis haben könnte.

Der übergreifende Beitrag dieser Dissertation ist ein neuartiger konzeptioneller Rahmen für CI-unterstützte Service-Interaktionen, der auf grundlegenden Theorien der menschlichen Kommunikation, der technologievermittelten Servicebereitstellung und auf laufenden Arbeiten zu Mensch-KI-Beziehungen aufbaut und diese integriert. Der konzeptionelle Rahmen wird in sechs komplementären Projekten entwickelt und getestet. In Kombination zeigen diese Arbeiten, wie CIs und ihr Design (z.B. das Geschlecht der Stimme, die Interaktionsmodalität) die Nutzererfahrungen (z.B. Vertrauen, Flow, Fluency) prägen und sich letztlich auf das Ergebnis der Dienstleistung auswirken. Darüber hinaus wird untersucht, wie moderierende Faktoren, die sich in der konversationalen Gestaltung von CIs manifestieren, diese Effekte beeinflussen. Die Dissertation hat wichtige Implikationen für Forschung und Praxis: Aus theoretischer Sicht zeigt diese Arbeit, wie grundlegende Erkenntnisse aus der menschlichen Kommunikation genutzt werden können, um die aufkommende interdisziplinäre Forschung zu dem Thema voranzutreiben, wie Nutzer in Beziehung zu den Service-Interaktionen von CIs treten und von diesen beeinflusst werden. Die Ergebnisse dieser Dissertation haben ebenfalls wich-

tige praktische Auswirkungen für Unternehmen, die das Potenzial von CIs zur Verbesserung der Nutzererfahrungen und das theoriegeleiteten Konversationsdesign von CIs nutzen wollen.

Stichworte

Sprachbasierte Interfaces, Sprachinteraktion, Sprachassistent, Benutzererfahrung, Konversationsdesign, Serviceergebnisse, Dienstleistungsforschung

Part A: Synopsis

1 Introduction

1.1 Problem Statement

Recent advances in artificial intelligence (AI) and natural language processing (NLP) have supported the proliferation of conversational interfaces (CIs)—also known as conversational agents (CAs) or smart personal assistants (SPAs)—that fundamentally change the ways in which we access and use services (Huang & Rust, 2018; Schuetz & Venkatesh, 2020). These interfaces have emerged as a novel interaction paradigm allowing firms to engage users in increasingly personal ways during self-service interactions, almost as though they were interacting with human service providers (Diederich et al., 2022; Hollebeek et al., 2021). CIs are an emerging class of self-service technology that allow users to engage in text- or voice-based dialogical interactions with service providers (Knote et al., 2021; Pfeuffer, Benlian, et al., 2019). From contesting parking tickets using the DoNotPay lawyer chatbot (Zack, 2020) to conducting money transfers through conversation with Bank of America’s virtual voice assistant Erica (Fuscaldo, 2019), users are increasingly engaging with CIs for support in accessing services more naturally than the use of non-conversational technologies, such as traditional websites, allows (Davenport et al., 2020; Miao et al., 2022). Recent industry studies indicate that the number of CIs is rising exponentially (e.g., Amazon Alexa’s skills alone have increased to over 100,000 as of 2021; Laricchia, 2022a), demonstrating CIs’ potential for the future of service provision.

CIs promise a wide range of benefits to both users and firms (Larivière et al., 2017; Wirtz et al., 2018). For users, they represent a novel service channel that provides personalized, round-the-clock customer support that is more immediate and convenient than that provided by traditional self-service channels (de Keyser et al., 2019). Moreover, CIs are believed to enhance the sense of “human interaction” when interacting with an automated computer interface, ultimately creating a more experiential service experience (Pfeuffer, Benlian, et al., 2019; van Doorn et al., 2017). A recent industry report indicates that over 50% of users prefer CIs over human service providers for help with more routine tasks, with this preference reaching 70% among those aged 18–34 years (Press, 2019), hinting at shifting user preferences with regards to service delivery processes.

For firms, CIs are important tools that can help enhance service effectiveness and efficiency (Hollebeek et al., 2021). Following the logic of cognitive automation (Engel et al., 2022), these interfaces enable companies to intelligently shift customer processes increasingly toward self-service leading to immediate benefits. First, CIs are expected to provide firms with up to USD 11 billion in annual savings by 2023 (Smith, 2018a). Second, the use of CIs enables companies to deploy their employees for more value-adding activities (e.g., a recent study by Salesforces found that 64% of service frontline employees report having more time for complex problems; Sweezey, 2019). Finally, these interfaces allow the collection of rich interaction data (e.g., vocal data) that can be used to perform advanced user analytics and act on these insights (Hildebrand et al., 2020). For instance, CIs can potentially perceive when a user is bored or happy in real-time and act to cease or increase engagement on this basis (Thomaz et al., 2020).

In light of these expected benefits, firms are increasingly adopting CIs to augment and automate selected service encounters across a wide range of industries and contexts, including retail (Qiu & Benbasat, 2009), health (Sezgin et al., 2020), and education (Winkler & Söllner, 2018). However, anecdotal evidence suggests that their effectiveness varies across firms (Miao et al., 2022). For instance, Clara, a chatbot used by the Swiss insurer Helvetia, successfully supports users in filing insurance claims (Hundertmark, 2021); however, IKEA was obliged to discontinue its chatbot, Anna, following numerous user complaints about its usability (Brandtzaeg & Følstad, 2018; Miao et al., 2022). Indeed, despite users' growing interest in the novel service channel, many CIs offer unsatisfactory service experiences, leading to ineffective and inefficient use or total discontinuation (Griffeth & Simonite, 2019).

At the same time, it is difficult to apply the insights gained from the extensive knowledge base on automated, non-conversational interfaces in service contexts to the effective design of CIs, as they possess unique features (Huang & Rust, 2021; Maedche et al., 2019). CIs differ from non-conversational interfaces concerning the mode of interaction and their appearance (Hildebrand & Bergner, 2020). First, CIs mimic human conversations by following a dynamic and to a varying degree adaptive interaction protocol in which the user and the interface take turns (Miao et al., 2022). Second, CIs afford to integrate a wide range of anthropomorphic cues, such as imbuing an interface with a voice that can enhance the human-likeness of their appearance (Feine et al., 2019). Thus, CIs enter spheres of communication previously reserved for humans: natural-language-based dialogue (Jurafsky & Martin, 2017). This capability may fundamentally affect how users respond to automated interfaces in service contexts. Hence, this dissertation

explores how CIs and their design affect the service experience, how they perceive the firm providing such conversational exchanges, and what underlying psychological mechanisms might explain these effects and their downstream consequences for users and firms.

1.2 Research Questions and Research Methods

This dissertation's overall goal is to investigate how CIs and their design affect user experiences and service outcomes, and its findings may be divided into two parts. The first part is conceptual in nature and aims to provide an integrative framework for CIs in service contexts (RQ1). Building on this foundation, the second part of this dissertation is empirically driven and addresses important characteristics of CIs, pertaining both to the interface (i.e., voice gender; RQ2) and the conversational interaction (i.e., modality; RQ3). Figure 1 provides an overview of the research questions (RQs).

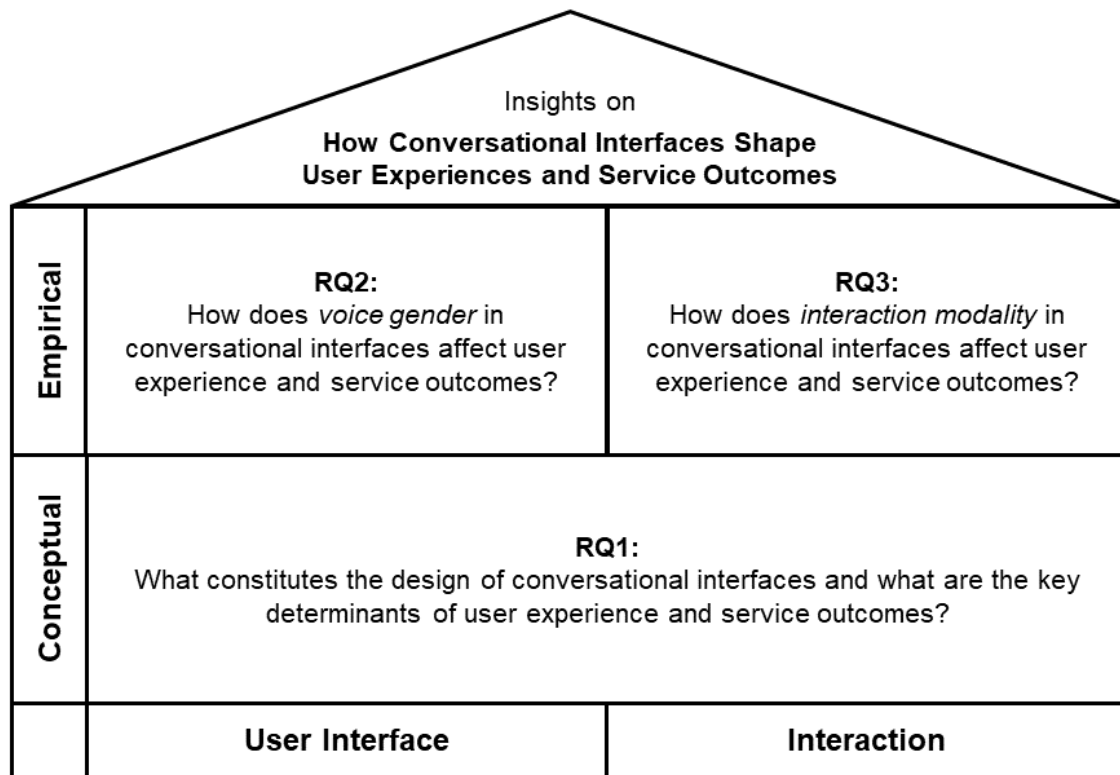


Figure 1. Overview of research questions

As CIs increasingly enter many people's lives and change the ways in which we interact with information systems, research interest in those interfaces has grown considerably. However, the abundance of literature makes it difficult to meaningfully integrate the growing knowledge base on CIs, not least as the literature on CIs is dispersed across various different disciplines, such as human–computer interaction (HCI), marketing,

and information systems (IS). Moreover, most studies examining and designing CIs have focused on individual or limited numbers of design elements or configurations and their effects on selected user perceptions and behaviors (Diederich et al., 2022). The existing research remains ambiguous with respect to defining and classifying key features of CIs and respective user outcomes, making it difficult for researchers to compare empirical results and draw meaningful conclusions for interface design and for the deployment of CIs along the service value chain (Miao et al., 2022) (see also Paper A & B).

RQ1: *What constitutes the design of conversational interfaces and what are the key determinants of user experience and service outcomes?*

This dissertation includes a series of systematic reviews of the current state-of-the-art on CIs to answer this question. Paper A presents the results of a systematic literature review (SLR) to summarize and synthesize the empirical literature on CIs. The results reveal five dimensions of interface characteristics (auditory, interaction, invisible, visual, and verbal cues), five dimensions of user outcomes (rapport, social presence, trust, utility, and behavioral intention and outcomes), and the respective strengths of the relationships between the interface characteristics and user outcomes. Paper B presents the results of a SLR to summarize and synthesize the literature on trust in CIs. The results reveal three distinct properties of CIs (i.e., anthropomorphism, dialogic interactions, probabilistic interactions) that uniquely influence how trustworthy user experiences are formed. As such, Paper A and B provide a comprehensive framework of CIs in service contexts.

Building on the conceptual foundation laid down in the first part of this dissertation, the second part aims to empirically investigate how key features of these interfaces affect user perceptions and service outcomes. In this regard, two main research perspectives can be distinguished: user interface-driven and interaction-driven experiences and behaviors. First, user interface-driven experiences and outcomes emerge from the static characteristics, such as audio-visuals. Second, interaction-driven experiences and outcomes arise from the dynamic dialogue between user and interface (Zierau, Engel, et al., 2020). As such, this dissertation assumes a holistic perspective on conversational interactions covering both the entity that communicates the message as well as the way how the message is communicated. In contrast, this dissertation does not separately address

the probabilistic nature of CIs. In the following, specific research problems relating to both perspectives are outlined:

The first specific issue that this dissertation addresses concerns the use of gender-specific anthropomorphic cues in voice-based CIs. Most voice-based CIs are designed to personify female characters. A recent study has shown that an overwhelming proportion—up to 77%—of all CIs use female characteristics (Feine et al., 2020). Indeed, most CIs produced by major technology providers, such as Amazon (Alexa), Apple (Siri), or Microsoft (Cortana) use female voices (as a default) (Hefti, 2022). This burgeoning trend raises the question of how users react to gendered voices in CIs and how they perceive the firm providing those service interfaces.

RQ2: *How does voice gender in conversational interfaces affect user experience and service outcomes?*

According to literature, there are gendered preconceptions about which voices sound the most trustworthy (Schmitt, Zierau, et al., 2021). Specifically, most research shows that people trust female voices over male voices (McAleer et al., 2014). There is also emerging research on the effect of gendered synthetic voices on user experiences, mainly showing that humans make trust attributions toward the interface based on voice gender depending on the task context (e.g., Nass et al., 1997). Building on this research, this dissertation addresses three specific research gaps. First, although previous literature shows that the effect of voice gender is highly contextually, no research investigates how the properties of the conversational interaction itself impact the effect of voice gender on interface trust. Second, there are few insights on the downstream impact of gendered voice-based interfaces on the firm they represent. Third, previous research exclusively operationalized voice gender using a binary conceptualization of gender. At the same time, technology firms are experimenting with gender-ambiguous voices to provide more inclusive user experiences (Sutton, 2020) (see also Paper C & D).

This dissertation presents the results of a series of experimental studies designed to fill these gaps. Paper C systematically investigates the combined effect of voice gender (female vs. male) and communication style (task-oriented vs. social-oriented) on interface trust and firm perceptions. The results illuminate a novel boundary condition of the effect of voice gender on interface trust (i.e., conversational style). Specifically, they show that a social-oriented style leads to reduced trust in male voices while a task-oriented style elicits no significant differences. Moreover, the results show that greater levels of

interface trust led to a more positively-valenced delivery experience and attributions of benevolence to the firm. Finally, paper D investigates the effects of a gender-ambiguous voice compared to gendered voices on interface trust and stereotype attribution across different tasks. Overall, the effects hint at the potential of gender-ambiguous voices as this study shows no significant differences in trait and trust attribution between gendered and non-gendered voices. These insights support an effective and inclusive voice design of CIs in service contexts.

The second issue that this dissertation addresses concerns interaction modality (voice vs. text). While text-based CIs (i.e., chat bots), have dominated customer service in recent years, CIs (i.e., voice bots) have recently gained prominence (Schmitt, Zierau, et al., 2021; Seaborn et al., 2021). Indeed, the number of voice-based CIs has grown considerably in recent years and is expected to reach 8.4 billion by 2024 (Smith, 2018b). This development raises the question as to how users react to voice-based CIs in service contexts as well as to the firms providing such voice-mediated automated service exchanges, the psychological mechanisms that underlie these effects, and the downstream consequences for service outcomes.

RQ3: *How does interaction modality in conversational interfaces affect user experience and service outcomes?*

Literature shows that voice-based CIs are associated with a more intuitive, positively-valenced delivery experience (Rzepka et al., 2021) that can impact consumer decision-making (Klesse et al., 2015) and the type of information users reveal about themselves (J. A. Berger et al., 2021). The findings illustrate how the specific characteristics of voice-based CIs (i.e., verbal cues, channel synchronicity; see Paper E) create unique interface experiences. Building on this research, this dissertation aims to address four specific research gaps. First, previous research has investigated voice-based modalities by conceptualizing voice as an input or an output. Few studies investigate a fully conversational voice-based service interaction. Second, no research investigates psychological mechanisms related to the unique affordance of a voice-mediated conversational interaction. Third, most research focused on boundary conditions manifested in interface appearance (e.g., avatar; Qiu & Benbasat, 2009) or user characteristics (e.g., privacy concerns; Pagani et al., 2019). However, no prior work examined the influence of the unique conversational design characteristics during an interaction. Finally, most research has focused on user perceptions of the interface (Qiu & Benbasat, 2005) rather

than investigating firm perceptions or downstream behavioral consequences (see also Paper E & F).

This dissertation presents a series of experimental studies designed to examine how the unique affordances of voice-based (as opposed to text-based) exchanges impact user experiences and service outcomes. Paper E demonstrates across four studies that voice-based CIs promote a more flow-like user experience, leading to a more positive-valenced delivery experience and more positive behavioral service outcomes (e.g., contract renewal). Moreover, Paper E provides evidence for two boundary conditions manifested in the conversational style that reduce flow in voice-based CIs (i.e., semantic disfluency and the number of conversational turns). Paper F demonstrates across two studies that voice-based CIs decrease processing fluency, leading to a less positive recommendation evaluation, ultimately reducing purchasing intention. Moreover, Paper F provides evidence for a condition that increases fluency in voice-based CIs (i.e., numerical framing of information). These findings have important implications for firms that aim to deploy voice-based CIs and effective conversational design.

1.3 Structure of this Dissertation

Following the guidelines regarding cumulative dissertations, this dissertation is broadly divided into two major parts: Part A consists of a synopsis paper, and Part B comprises six individual research papers that are independent but related and which follow the structure of the research program sequentially as it is laid down in this dissertation.

The synopsis introduces and positions the dissertation's main results and discusses their contribution against the background of the dissertation's overall goal. Chapter 2 introduces and discusses the core terms of this dissertation and develops a conceptual framework within which the dissertation's results can be positioned. In Chapter 3, the fundamental tenets of the research strategy implemented in this dissertation are explained and discussed. Chapter 4 presents the main results across each publication. Finally, in Chapter 5, the dissertation's contributions to theory and practice and its limitations are discussed, and avenues for future studies are highlighted to build and expand on these findings and advance research on CIs in service contexts.

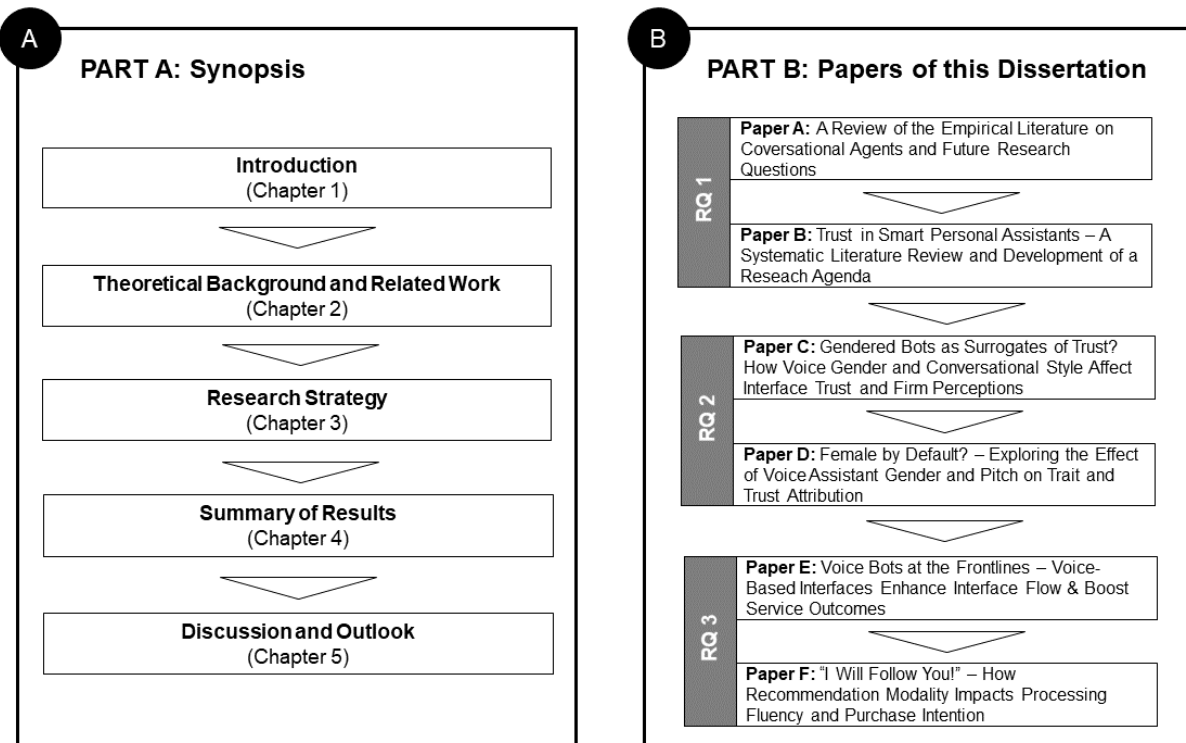


Figure 2. Structure of this dissertation

2 Theoretical Background and Related Work

This chapter's purpose is to provide an overview of this dissertation's major concepts of this dissertation and to develop a general conceptual framework within which the dissertation's findings can be placed. First, a brief overview of the history of CIs from the first technological developments to their widespread use in customer service is provided. Next, differences from other service encounter types and the individual properties of dialogical service encounters are discussed. Basic communication theories are then explored as a basis for this dissertation's theorizing. Moreover, the respective user experience dimensions investigated in this dissertation are presented. Finally, service outcomes and the dissertation's basic conceptual framework will be briefly discussed.

2.1 The Past and Present of Conversational Interfaces

What has been the developmental trajectory of CIs providing customer service to humans? Based on the accounts of Kontogiorgos (2022) and Adamopoulou & Moussiades (2020), a brief history of the development of CIs is given below, and their key features are highlighted. The origins of conversational technologies can be traced to the first half of the twentieth century. In 1938, the Bell Telephone Laboratory developed a synthetic speaker called Voder (Voice Operating Demonstrator). This machine was controlled by a human operator via a keyboard. By pressing different keys, the operator could control various voice characteristics, such as frequency, via a complex mechanical mechanism. The operation of Voder required one year's training, testifying to these early machines' complexity. While Voder could not automatically interact with users, it formed the basis for synthesized speech as we know it today (Dudley et al., 1939; Kontogiorgos, 2022).

With the advent of the first programmable digital computers in 1943 and 1945, scientists focused increasingly on the question of how machines could communicate with humans naturally and intelligently (Kontogiorgos, 2022). In 1950, Alan Turing at the University of Manchester developed the Turing test, which he called the "imitation game", to identify whether machines can be considered intelligent. According to the Turing test, a machine may be regarded as intelligent when a human interacts with both a human and the computer through a keyboard and cannot distinguish between the two (Kontogiorgos, 2022; Turing, 1950). In 1964, Joseph Weizenbaum developed a computer program called ELIZA at the Massachusetts Institute of Technology that was expected to be the first to pass the Turing test. ELIZA was designed to simulate text-based conversations

following a rule-based interaction logic: it could process simple human text input and respond with scripted answers (Kontogiorgos, 2022; Weizenbaum, 1966).

Over the next 20 years, computer agents' ability to interact with humans was successively improved and extended to other modalities, such as voice. The following are just a sample of these critical developments (Kontogiorgos, 2022). In 1972, Terry Winograd at the Massachusetts Institute of Technology developed *SHRDLU*, a computer program that was designed to interact with humans using NLP to manipulate simulated building blocks in a virtual environment (Kontogiorgos, 2022; Winograd, 1972). In the same year, WABOT-1, an anthropomorphic robot, was developed at Japan's Waseda University. WABOT-1 could communicate with users in very basic Japanese (Kato et al., 1972; Kontogiorgos, 2022). This was followed by PARRY, a computer program that was developed in 1972 by Kenneth Colby at Stanford University and that was supposed to be more advanced than ELIZA, as the developers aimed to imbue the computer agent with a "personality". As such, PARRY could provide different emotional responses as the valence of users' inputs changed (Colby et al., 1971; Kontogiorgos, 2022). In 1980, Richard Bolt at the Massachusetts Institute of Technology developed the "put-that-there" system, an initial multi-modal interface that combines speech and gesture recognition to control a graphical user interface (Bolt, 1980; Kontogiorgos, 2022). In 1979, the first major commercial product using speech synthetization was developed: *Speak & Spell* was designed by Texas Instruments as a hand-held computer to support children in learning pronunciation. The children could enter letters and words, which were then reproduced by the embedded speech synthesizer (Frantz & Wiggins, 2013; Kontogiorgos, 2022).

Technological developments accelerated during the 1990s, leveraging and combining the growing knowledge base in AI, NLP, and the increasing proliferation of the Internet (Adamopoulou & Moussiades, 2020). In 1991, TinyMUD was developed by Michael Mauldin at Carnegie Mellon University as an artificial computer player designed to chat with users. Interestingly, it emerged that users preferred to talk to TinyMUD than to other human players. The term "chatterbot" first appeared in relation to the development of TinyMUD (Adamopoulou & Moussiades, 2020; Mauldin, 1994). In 1995, ALICE was designed as the first CI that could be accessed via the internet. While ALICE still followed a rule-based approach, it was already possible to converse with ALICE about various topics over longer periods of time (Adamopoulou & Moussiades, 2020; Wallace, 2009). Finally, in 1997, IBM developed Deep Blue, an AI-based chess-playing computer program that was capable of defeating Garry Kasparov, the reigning world

chess champion at the time, testifying to AI's growing potential to become more efficient than humans at certain tasks (Adamopoulou & Moussiades, 2020; Campbell et al., 2002).

Further technological advances in NLP and AI, the proliferation of increasingly powerful devices (e.g., smart phones, smart speakers), ultimately led to the commercial breakthrough of CIs in the 2000s (Adamopoulou & Moussiades, 2020). Starting in 2001, SmartChild was developed as a text-based CI, which could be accessed through messaging services such as MSN as a means of accessing support in simple information tasks, such as getting the latest stock prices (Adamopoulou & Moussiades, 2020; S. Singh & Thakur, 2020). Successively, large technology providers starting with Apple in 2011 (Siri) and Amazon in 2014 (Alexa) developed intelligent personal assistants that support the user with an increasing number of tasks such as setting timers, providing weather information or order products across modalities. These interfaces have little in common with early computer programs such as ELIZA but can now interact adaptively and personally with the user and learn to do so even better with each interaction (Adamopoulou & Moussiades, 2020). Furthermore, the emergence of standardized conversational technology platforms and low-code programming environments allowed developers and laymen to easily develop and deploy CIs. This has led to a surge in CIs across various domains from healthcare to customer service (Winkler et al., 2020).

Today, the growing proliferation of CIs is gradually but steadily changing how consumers access and use self-service (Amazon, 2015; Smith, 2019). CIs are powered by recent computational advancements that combine both NLP and AI to facilitate a fully conversational automated service interaction (Knote et al., 2021; Pfeuffer, Benlian, et al., 2019). Examples of CIs in service encounters are abound, ranging from general purpose interfaces allowing users to perform a plethora of mundane assisted tasks, such as booking flights, reading the news, or adding items to shopping lists, to highly specified interfaces that provide expert advice on insurance products or investment strategy (Thomaz et al., 2020). Based on the customer service context, those interactions either take the form of a one-off engagement or are part of an ongoing relationship with an existing customer (Miao et al., 2022). As such, CIs are becoming omnipresent at service frontlines (Wirtz et al., 2018). This burgeoning trend toward the deployment of CIs raises the questions of how users respond to these interfaces and their design in service settings, how they perceive service firms that provide such automated conversational interactions, and the psychological mechanisms that underlie these effects as well as the downstream consequences for users and organizations.

2.2 Distinguishing Conversational Interfaces From Other Self-Service Channels

Why do CIs require novel research efforts? As Table 1 shows, the proliferation of CIs has witnessed a fundamental shift with respect to both (1) the entities used in service systems and (2) the ways in which service is delivered (Komiak & Benbasat, 2003). The initial shift was from interpersonal to website-mediated service encounters. Website use has imposed several spatial and temporal separations between service providers and users (Gefen & Straub, 2003), and computer interfaces are becoming the primary service channel for many firms. These interfaces represent the company in the outside world and have become the focal crystallization point of the overall service experience (Hildebrand & Bergner, 2020). Recently, the emergence of CIs has yielded another critical paradigm shift from static interaction with websites toward communication with an “intelligent” computer interface via natural language (Maedche et al., 2019). In short, CIs have emerged as a novel service interaction paradigm, allowing service firms to engage users in increasingly personal ways, almost as though they were talking to a human service employee, thus bridging the gap between interpersonal and website-facilitated service encounters (Hollebeek et al., 2021; Larivière et al., 2017).

Table 1. Differences among prevalent service encounters (adapted from Komiak & Benbasat, 2003)

	Interpersonal	Website	Conversational Interface
Entities	Employee	Website	Conversational Interface
	Service Provider	Information on service provider on a website	Information on service provider facilitated by conversational interface
	Service (facilitated by employee)	Service (facilitated by website)	Service (facilitated by conversational interface)
Service Delivery	Dialogic (Human-to-human)	Static (Human-to-computer)	Dialogic (Human-to-computer)
	Probabilistic	Deterministic	Probabilistic

The literature indicates that these CIs represent a paradigm shift in human–computer interaction, fundamentally changing the ways in which services are delivered to users (Mende et al., 2019). From a sociotechnical perspective, CIs represent a novel form of IS that is distinguished by its high degree of interactivity and intelligence (Maedche et al., 2019). These capabilities may fundamentally affect user experiences and service outcomes and raise several theory- and design-related questions, most prominently revolving around an emergent conversational interaction paradigm (Hildebrand & Bergner, 2020) and the increasingly autonomous and “black box”-like character of AI-based technologies (Rahwan et al., 2019). Thus, substantial knowledge created on traditional self-service channels (i.e., websites) cannot necessarily be adapted to CIs (Rai, 2017). It is thus critical to understand how user experiences and service outcomes are impacted by CIs and their design (Paper A & B). This dissertation focuses on the dialogic interaction mode of CIs. In contrast, the probabilistic nature of CIs is not focused on in this dissertation.

2.3 The Unique Features of Conversation-Based Service Delivery

What makes dialogic interactions with CIs uniquely useful for service delivery? According to Hildebrand & Bergner (2020), several malleable features of the dialogue-based interaction mode of CIs ultimately enable a more interactive and personalized mode of service delivery than traditional self-service channels, such as websites (see Table 2). First, the interaction flow follows a dynamic protocol in which the user and the interface take turns to communicate with one another. Thus, service requests are no longer processed according to a simple input and output logic (e.g., as realized in static questionnaires) but rather imitate an ongoing dialogic interaction between the user and the service employee (Hildebrand & Bergner, 2020; Miao et al., 2022). As such, linguistic style has also evolved toward a more informal form of communication, including trivial acknowledgments to move the communication forward (Hildebrand & Bergner, 2020; Thomaz et al., 2020). Second, CIs permit the integration of a wide range of anthropomorphic cues (e.g., avatars) that can further enhance the sensation of “human interaction” in automated computer interfaces (Hildebrand & Bergner, 2020).

Table 2. Difference between website- and conversational interface-mediated service encounters (adapted from Hildebrand & Bergner, 2020).

	Website	Conversational Interface
Interaction Features	• One-way (Static) • Formal language • Absent anthropomorphic design cues	• Turn-taking (Dynamic) • Informal language • Present anthropomorphic design cues

In recent years, researchers have begun to investigate the effects of some of these malleable features on user experience and service outcomes, demonstrating that conversational interactions can enhance user trust in the interface with positive downstream effects for firm perceptions and investment behavior (Hildebrand & Bergner, 2020) and that anthropomorphizing the interface can positively affect a wide range of outcomes, including trust (Qiu & Benbasat, 2009) and service perceptions (A. M. Seeger et al., 2021). The first part of this dissertation attempts to synthesize and integrate this recent work on CIs in service contexts (RQ1). Moreover, this dissertation aims to investigate two key features of CIs. First, contributing to the literature on the anthropomorphization of interfaces, the effects of gendered voices in CIs are systematically explored (RQ2). Second, by investigating the effect of interaction modality (voice vs. text), this dissertation contributes to the scientific discourse on how conversational interactions with automated service interfaces affect user experience and service outcomes (RQ3). A brief outline of research on each of these design aspects is provided below.

2.3.1 Feature 1: Voice Gender

Most firms utilize female voices in CIs (Feine et al., 2020; Hefti, 2022). This burgeoning trend raises the question of how users respond to gendered voice-based CIs with which downstream consequences for the firms they represent. In interpersonal interactions, there are gendered preconceptions about the trustworthiness of voices (McAleer et al., 2014). Accordingly, one primary source of trust attributions towards voices are social heuristics based on gender stereotyping. Cross-culturally, gender is a significant factor in personal stereotyping (Eckes, 2008). According to the content stereotype model, many people associate gender with specific traits and behaviors: women are usually regarded as warm and friendly, while men are perceived as competent and capable (Fiske

et al., 2002). These trait interferences influence the way how people attribute trust to each other (R. C. Mayer et al., 1995). Research has shown that people are likely to attribute the same characteristics to computer interfaces (Moon, 2000). In the context of HCI, research has shown that gendered voices alone can lead a user to label a computer to be female or male and to assign stereotyped roles to them (Tay et al., 2014).

In the context of computer interfaces, initial findings suggest that voice gender can indeed influence trust attributions depending on the context (Nass et al., 1997). However, as Paper C & D highlight, several gaps remain concerning the effect of gendered voices. First, most studies implemented scenario-based approaches that featured no dynamic interaction between the interface and the user (e.g., Tay et al., 2014). Few studies operationalized fully conversational interactions to study trait and trust attribution toward voice-based CIs. Second, prior research has shown that the influence of voice gender on trust depends on the application context (Nass et al., 1997). Existing research has mainly investigated boundary conditions relating to task context (e.g., influence of the interaction with stereotyped tasks; Eyssel & Hegel, 2012). No prior work has explored the unique properties of conversation design during an interaction, such as the degree of formality. Third, the outcome types studied focused predominantly on trust perceptions toward the interface (e.g., A. Kim et al., 2019). However, whether gender-specific cues unique to voice-based interaction shape firm-related perceptions remains unclear. Finally, all studies used a binary gender conceptualization (female vs. male; e.g., Nunamaker et al., 2011). At the same time, companies increasingly experiment with gender-ambiguous voices to provide more inclusive user experiences (see Paper C & D).

Hence, this dissertation examines fully conversational service experiences using gendered and gender-ambiguous voices and aspects of conversational design (i.e., communication design) and captures perceptual outcomes relevant to both the user and the firm.

2.3.2 Feature 2: Interaction Modality

Researchers largely distinguish between text-based and voice-based CIs (Diederich et al., 2022). Voice-based CIs differ from text-based interaction modalities with respect to both cue and channel-level characteristics (see also Paper E for a comprehensive conceptualization of the affordances of voice-based as opposed to text-based CIs). First, considering cue-level characteristics, verbal cues encompass the vocal features of language (e.g., pitch, accent), while textual cues refer to written language (Moffett et al.,

2021). Second, considering channel-level characteristics, voice-based interactions are more synchronous than text-based interactions. Moreover, contrary to text-based interactions, voice-based CIs provide low levels of channel revisability as they do not allow users to re-read or re-type messages (Moffett et al., 2021). In short, voice-based CIs provide unique affordances (i.e., verbal cues, channel synchronicity) compared to text-based interfaces, which can fundamentally affect user experiences and service outcomes (Paper E).

In recent years, researchers have begun to investigate the malleable features of voice-based (as opposed to text-based) CIs and their effects on selected user perceptions and service outcomes, demonstrating that voice-based CIs can create a more positive user experience (Rzepka et al., 2021) capable of altering consumer preferences (Klesse et al., 2015) and increasing information disclosure (J. A. Berger et al., 2021). However, as Paper E & F highlight, several gaps remain. First, the majority of work to date has assessed voice-based modalities by operationalizing voice as an input (e.g., Klesse et al., 2015) or an output (e.g., Qiu & Benbasat, 2009). Few studies operationalized voice-based CIs as fully conversational interactions requiring voice as both an input and an output (Rzepka et al., 2021). Second, existing research has mainly investigated boundary conditions relating to the interface's appearance (e.g., influence of avatars; Hess et al., 2009) or user characteristics (e.g., privacy concerns; Pagani et al., 2019). No studies have investigated the unique aspects of conversational design (i.e., semantics, structure). Third, the psychological mechanisms studied focus primarily on perceived social presence (Qiu & Benbasat, 2009). However, it remains unclear whether the interactivity-related affordances of voice-based CIs also change how users experience service delivery. Finally, the service outcome types studied consist predominantly of perceptual outcomes (e.g., interface trust; Qiu & Benbasat, 2009). Meanwhile, behavioral outcomes or firm perceptions have not been considered hitherto (see Paper E & F).

Hence, this dissertation examines fully conversational service experiences, key conversational design characteristics (semantics, structure, framing), two interactivity-related mechanisms (flow, fluency), and perceptual and behavioral outcomes relevant to the service interaction.

2.4 Communication Theories

The theorizing in this dissertation mainly builds on prevalent theories on (human) communication. Hence, in this section, some of the most widely accepted communication

theories are reviewed, selected for their applicability to the focal topics of this dissertation (social response theory, media richness theory). They constitute the primary theoretical lenses that are used in this dissertation to hypothesize on the effect of different CIs and their design on user experience and service outcomes.

2.4.1 Social Response Theory

Social response theory —also called the computer are social actors (CASA) paradigm—posits that social cues in nonhuman technologies, such as automatic service interfaces, can trigger social responses (Moon, 2000; Nass & Moon, 2000). According to social response theory, humans tend to apply social categorization and heuristics to computers, a process known as anthropomorphism (Moon, 2000). As such, various social cues, including visual, auditory, and verbal cues, can increase users' tendencies to approach the interface as a relevant social entity (Moffett et al., 2021; Nass et al., 1994; A. M. Seeger et al., 2021). This tendency to anthropomorphize computers is deeply rooted in human psychology. Accordingly, anthropomorphism can both satisfy the need for control and social connection (Epley et al., 2008). As such, research has shown that anthropomorphism can affect a host of different user perceptions, such as likability (T. W. Bickmore et al., 2005) and trust (A.-M. Seeger & Heinzl, 2021) and behaviors such as information disclosure (Moffett et al., 2021; Wambsganss et al., 2020). In this dissertation, social response theory is implemented as the underlying theory to explain the effects of voice gender on user experiences and service outcomes.

2.4.2 Media Richness Theory

According to media richness theory, computer interfaces can differ in the degree of information richness that they provide to users. Information richness refers to the rate at which information can be exchanged within a given time frame (Daft & Lengel, 1986). Several cues and channel-related characteristics enhance media richness in interfaces (Moffett et al., 2021). First, interfaces can provide various visual (e.g., avatars), auditory (e.g., voice qualities), and verbal (e.g., communication style) cues that enhance media richness (Dennis & Kinney, 1998; Moffett et al., 2021). Second, media richness is strongly linked to the degree of channel synchronicity that the interface provides to the user (Daft & Lengel, 1986; Moffett et al., 2021). Interfaces that provide a mode of interaction that is high in synchronicity enhance media richness (Hoffman & Novak, 1996). Although, media-rich (as opposed to media-lean) service interfaces are associated with increased communication costs (e.g., cognitive load), researchers identified

various beneficial user outcomes (Moffett et al., 2021). Thus, media richness is associated not only with higher vividness and interactivity and its downstream effects, such as enjoyment (van Kerrebroeck et al., 2017), but also with the increased trustworthiness of the information exchanged (Moffett et al., 2021; Rockmann & Northcraft, 2008). In this dissertation, media richness theory is mainly used as a basis to theorize about the unique properties of voice-based (as opposed) to text-based CIs.

2.5 User Experience

A key aim of this dissertation is to investigate how CIs and their design impact user experiences. According to the literature, user experience is a multi-faceted phenomenon integrating a multitude of experience dimensions (Følstad & Bae, 2020; Hassenzahl & Tractinsky, 2006; Law et al., 2009). This dissertation focuses on three distinct user experience dimensions based on the unique characteristics of CIs:

First, this dissertation addresses risk-related perceptions, such as user trust. For many users, CIs represent a novel kind of information system with which they have little experience. As such, de Bellis & Venkataramani Johar (2020), when comparing different adoption factors, found that a lack of trust in CIs in e-commerce is among the key barriers to use. This sentiment was also reflected in a Forbes study showing that 41.5 percent of U.S. consumers do not trust any AI-infused digital assistants (Kroque, 2017). Hence, users need to reduce the perceived risk associated with using CIs, and trust has been found to be a useful mechanism in reducing risk in IS across multiple contexts (Söllner, Benbasat, et al., 2016). Second, this dissertation investigates interactivity-related perceptions, such as flow and processing fluency. The existing literature demonstrates that interface activity can lead to cognitively engaging, enjoyable user experiences (Agarwal & Karahanna, 2000). At the same time, highly interactive interface can increase to high cognitive load, which can be burdensome in some application contexts (Chandra et al., 2012; Moffett et al., 2021). Therefore, this dissertation investigates both positively valenced (i.e., flow) and negatively-valenced (i.e., processing fluency) experience dimensions related to interactivity.

In sum, considering the unique characteristics of CIs (risk, interactivity), this dissertation focuses on trust, flow, and processing fluency as key interface-level variables that facilitate positive downstream consequences for users.

2.5.1 Trustworthy User Experiences

How do users cope with perceived risks associated with interface usage? In social interactions, trust is a mechanism that significantly influences people's behavior toward others and organizations (R. C. Mayer et al., 1995; Rousseau et al., 1998; Schroeder & Schroeder, 2018). Based on growing automation, more and more information systems are also becoming the target of such trust relationships (J. D. Lee & See, 2004; Söllner, Benbasat, et al., 2016). Nowadays, humans must rely on an automated interface that autonomously and independently processes the user request in many contexts (Söllner et al., 2012). As a result, trust is becoming increasingly crucial for the use of IS and is driving their adoption across a wide range of contexts (McKnight et al., 2011; Qiu & Benbasat, 2009; Söllner, Benbasat, et al., 2016). In the context of this dissertation, trust is defined as a *“psychological state comprising the intention to accept vulnerability based upon positive expectation of the intentions or behavior of another [i.e., a CI]”* (Rousseau et al., 1998, p. 395). Moreover, trust not only arises from the relationship with the IS itself but is also influenced by the salient entities of an application context. For instance, in the service context, especially the trust in the interface can interact with the trust in the service provider (Söllner, Hoffmann, et al., 2016). In this dissertation, trust is used to study the impact of voice gender on user experience and firm perceptions (see Paper C & D).

2.5.2 Absorbing User Experiences

How do the dialogic properties of CIs impact user perceptions and behaviors? An important stream of literature associates interface interactivity with immersive user experiences (Trevino & Webster, 1992). Flow is a concept that effectively describes such immersive experiences in computer-mediated interactions (Agarwal & Karahanna, 2000; Hoffman & Novak, 1996). Flow can be defined as state of deep immersion, focus, and enjoyment when interacting with an information system, such as a CI (Csikszentmihalyi, 1975). Flow has been shown to impact various user outcomes, including a more enjoyable interaction experience and a feeling of deep absorption when interacting with an information system (Trevino & Webster, 1992). Existing research has demonstrated that IS' individual characteristics, such as modality, can enhance or diminish flow experiences (D. Kim & Ko, 2019). Specifically, flow results when the interface permits a *“seamless sequence of responses facilitated by machine interactivity”* (Hoffman & Novak, 1996, p. 57). In this regard, the interface's interactivity and vividness are known to

enhance flow (Hoffman & Novak, 1996). Moreover, the interface must provide an optimal experience, balancing skill and challenge in a state in which the user is generally experiencing some degree of challenge (Clarke & Haworth, 1994). To that end, flow can help elucidate how users navigate conversational interactions. In this dissertation, the flow construct is used to study the impact of interaction modality on user perceptions and behavior (see Paper E).

2.5.3 Fluent User Experiences

On the other hand, interface interactivity is also associated with the efficiency of information exchange (Moffett et al., 2021). Specifically, it is assumed that interface interactivity influences message intelligibility (Dennis et al., 2008). An important concept used to study intelligible user experiences and their downstream consequences is processing fluency (Schwarz, 2004) and has been shown to affect a wide range of user perceptions and behaviors, including purchasing behavior (Morgan et al., 2021). According to existing research, various interface characteristics can affect processing fluency, including readability (Sung et al., 2022) and framing or presentation of information (Morgan et al., 2021). To understand how interface modality affects user experience and service outcomes in a product recommendation context, processing fluency is used as a key psychological mechanism to understand how users process recommendations across text-based and voice-based CIs (see Paper F).

2.6 Conversational Design

As argued above, CIs can be distinguished from other automated interfaces in their capability for natural-language-based dialogue (Jurafsky & Martin, 2017). Thus, CIs can use an extensive repertoire of conversational design elements to interact with users akin to human dialogue (Feine et al., 2019). According to previous research, how a conversational interaction is designed can critically influence how users experience CIs and their design (Pogacar et al., 2018). Thereby, conversational characteristics can be broadly distinguished into those that pertain to the content and process of communication (H. H. Clark & Brennan, 1991). Content characteristics refers to all elements that influence what is being conveyed (e.g., conversational style). On the other hand, the process characteristics refer to all elements influencing how the conversation is structured (e.g., the number of conversational turns) (H. H. Clark & Brennan, 1991; L. Clark et al., 2019; Pogacar et al., 2018). Together, these characteristics determine how the conversational interaction between the user and the interface unfolds. In the scope of

this dissertation, conversational design is used as a boundary condition to the effect of voice gender (Paper C & D) and interaction modality (Paper E & F) on user experiences and service outcomes.

2.7 Service Outcomes

Satisfactory user experiences with automated technologies are instrumental for a wide range of service outcomes (Hoffman & Novak, 1996). For instance, in the context of text-based CIs, Hildebrand & Bergner (2020) show that the conversational interaction with robo advisors can evoke a more trustworthy interface experience, ultimately influencing attributions of benevolence toward the firm and investor behavior. Within the scope of this dissertation, a wide variety of service outcomes are investigated to create a broad empirical basis for the use of CIs along the service value chain (Mende et al., 2019; van Doorn et al., 2017; Xiao & Kumar, 2021). On the one hand, critical perceptual variables are investigated that aim at assessing the overall service experience and the evaluation of the firm employing those CIs. On the other hand, this dissertation assesses behavioral outcomes consequential to service firms (e.g., conversion rate). Moreover, this dissertation aimed to investigate both outcomes that are immediate to CI-mediated user experiences and those that have extended effects on users and firms (i.e., contract renewal).

2.8 Conceptual Framework

Figure 3 illustrates the overarching conceptual framework¹ of this dissertation. In total, this framework stipulates three assumptions about how CIs and their design affect user experiences and service outcomes:

- *First*, the dissertation assumes that the unique characteristics of CIs (i.e., anthropomorphic appearance, conversational interaction) uniquely influence user experiences, specifically the trustworthiness, immersion and intelligibility of CIs. In particular, interface trustworthiness is assumed to be influenced by the anthropomorphic nature of the CI (e.g., voice gender). Moreover, conversational interactions and specifically interaction modality are assumed to influence the immersion and intelligibility of CIs.

¹ This conceptual framework is by no means to be seen as comprehensive nor exhaustive for the topic of service interactions with CIs but is formed based on the core themes of this dissertation.

- *Second*, this dissertation assumes that the unique properties of conversational design affect how CIs and their design influence user experiences. As such, conversational design is introduced as an important boundary condition to how users experience specific CI designs (e.g., voice gender, interaction modality)
- *Third*, this dissertation assumes that the user experience with CIs affects final service outcomes on both the perceptual and behavioral levels. Thus, the user experience with CIs can be seen as a crystallization point for the entire service experience.

These assumptions are tested over the scope of the six paper projects incorporated in this dissertation.

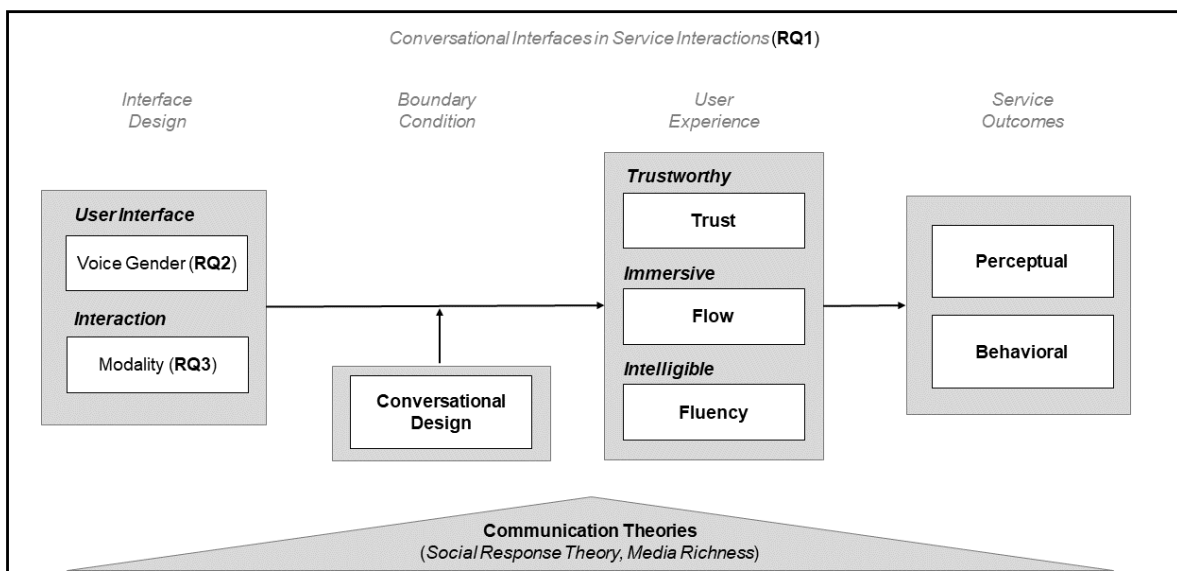


Figure 3. Conceptual framework of this dissertation

3 Research Strategy

This chapter outlines the philosophical and methodological considerations that form the foundation of the research strategy adopted in this dissertation. First, the epistemological background that underpins this research is elaborated. Second, the various research methods employed throughout this dissertation are discussed.

3.1 Epistemological Foundation

Epistemology is concerned with how knowledge can be defined and created (Döring & Bortz, 2016). Generally, three different strategies for knowledge creation can be differentiated (Recker, 2013). First, induction refers to the inference of general findings from a set of specific observations. More specifically, patterns and commonalities are identified in observed data points to develop novel theoretical propositions. This strategy underlies most qualitative research paradigms (Recker, 2013). Second, deductive knowledge creation strategies develop specific arguments or hypotheses based on a systematic synthesis of more general premises that are known from theory. These so-called “hypotheses” are tested using novel data points. In general, deduction mostly encompasses quantitative research approaches (Recker, 2013). Finally, abduction describes the process of finding the best possible explanation for a specific observation. As such, abduction involves exploratory, iterative research with the aim of pragmatically discovering new ideas. As opposed to induction and deduction, abduction *“is more concerned with discovery or design than rationalization or validation”* (Recker, 2013, p.42). Specifically, the design science paradigm may be regarded as a specific form of abductive research. Here, researchers aim to investigate and optimize the “usefulness” of certain interventions in real-world contexts (Döring & Bortz, 2016; Walton, 2014).

This dissertation draws primarily on a deductive knowledge creation strategy (Recker, 2013). Specifically, the research presented in this dissertation aims to investigate how CIs and their design affect specific user experiences and service outcomes and to shed light on the psychological mechanisms underlying these effects. While CIs represent a novel phenomenon in service contexts, fundamental theories of human communication (i.e., media richness, social response theory) can be applied to advance our understanding of how people perceive and respond to these novel interfaces. As such, these established and well-tested theories (as laid out in Chapter 2.4) form the basis for the deduction of hypotheses that are tested using behavioral experiments. In short, this disserta-

tion's main objective is to adapt and examine these fundamental theories of human communication in a new context (i.e., CIs in service encounters) and to generate novel insights and incremental theoretical advancements regarding the use of CIs in service contexts.

This dissertation contributes to a wide range of research disciplines spanning IS, marketing, and HCI. However, the research presented herein consistently coheres around the application of IS across different contexts and tasks, which is why the IS discipline is selected as the core research domain. IS research typically uses multi- and transdisciplinary research designs that span the social and natural sciences to investigate phenomena linked to IS artifacts (Engel, 2022; Recker, 2013). The social sciences focus on studying social processes that place the individual or groups of individuals (e.g., organizations, communities) at the center of scientific inquiry. For example, the systematic investigation of how users react toward CIs in service settings adopts this more behavioral-oriented perspective. Natural sciences are concerned with phenomena that manifest in material reality. To that end, IS research can assume an engineering perspective by exploring the design of IS artifacts. For instance, building and evaluating a CI that supports users in effectively and efficiently solve a service request based on theory-motivated design principles follows this more engineering-oriented approach to IS research (Engel, 2022; Gregor & Hevner, 2013). As described above, this dissertation is mainly anchored in social science within IS epistemology.

3.2 Research Methods

This section presents the methodological foundations of the research methods used to address the research questions. Further detail on their application is provided in the individual papers.

3.2.1 Systematic Literature Review

Research on CIs in service contexts is a nascent and emerging research field across several disciplines. Proliferated by the increased adoption of CIs across industries and tasks, research on these interfaces has grown exponentially in recent years (Diederich et al., 2022). To provide an integrated conceptualization and synthesis of research on CIs in service contexts, this dissertation relies on SLRs (Torraco, 2005).

SLRs incorporate a comprehensive set of techniques for the systematic identification, analysis, and synthesis of literature (vom Brocke et al., 2015; Webster & Watson, 2002). Hereby, the objective is often not only to review and summarize existing research but

also to derive novel theoretical insights and to develop an agenda for future research (vom Brocke et al., 2009). As the design and application of IS is propelled into different research disciplines, it is necessary to combine and build on diverse knowledge bases (Benbasat & Zmud, 2003; Engel, 2022). Specifically, the phenomenon of CIs represents a research field that spans different application scenarios and thus attracts the interest of multiple disciplines, among them IS, HCI, and marketing (Diederich et al., 2022; Knote et al., 2021). These disciplines provide valuable perspectives that support us in better understanding the effects of CIs and their design in service contexts. Hence, to account for this phenomenon's multi-disciplinary character, this dissertation draws on these and adjoining disciplines to provide an integrated perspective on CIs.

This dissertation aims to present the literature on CIs in service contexts transparently (i.e., systematically, explicitly, and reproducibly; vom Brocke et al., 2015). To that end, representative literature samples that account for the multi-disciplinary nature of service research were identified (Engel, 2022). Both outlet-based and database-based search strategies using a comprehensive set of keywords that best summarize the phenomena of interest were followed. Outlet-based strategies were applied when the object of interest was particularly broad (see Paper A for a review of the empirical literature on CIs), while database-based searches were used for more specific topics (see Paper B for a review of trust in CIs). In all cases, additional backward and forward searches were performed to identify relevant publications that were either cited in those articles identified in the initial search or that referenced them (vom Brocke et al., 2015). For the paper analysis, a concept-centric approach was employed and operationalized through the creation of concept matrices that emerged based on the systematic analysis of the retrieved papers (Webster & Watson, 2002). To that end, these matrices incorporate each publication focal scope of analysis (e.g., design lever, psychological mechanism, user perceptions and behaviors) and its contribution to theory and practice. Moreover, an approach based on that of Jeyaraj et al. (2006) (for further information, see Paper A) was followed in Paper A to identify areas in which consistent findings emerged having reached maturity and areas of inconsistent or few findings to identify future research opportunities.

3.2.2 Experimental Research

As outlined in Section 3.1., the research presented with respect to RQ2 and RQ3 builds on a quantitative paradigm. Specifically, several experimental studies were designed to investigate how CIs and their design influence user experience and service outcomes.

The methodology selected and the experimental paradigm used are described in greater detail below.

According to Mcgrath (1995), quantitative methods differ in their ability to provide (1) generalizability (i.e., degree of representativeness of the sample), (2) precision (i.e., degree of control and measurability), and (3) extential realism (i.e., degree of real-world applicability). In light of this dissertation's objectives, web experiments offer several distinct advantages (Recker, 2013): First, web experiments provide the tightly controlled setting (i.e., identical interaction and only variation in the type of interface used) necessary to observe nuanced psychological effects and their downstream consequences. Second, web experiments provided the opportunity to rapidly evaluate a range of different treatments (i.e., interface conditions), reaching high degrees of precision while achieving large sample sizes. To counteract the loss of realism often associated with experimental research, an attempt was made to make the design of the experimental platforms used as realistic as possible and to employ relatable service scenarios. Additional attempts were made to capture consequential behavioral variables (e.g., conversion rate; Paper E) to complement psychometric data.

To develop a rigorous and valid experimental paradigm, established guidelines were followed (e.g., Kirk, 1982; Recker, 2013): First, it was ensured that the experimental stimuli (i.e., interface conditions) differed sufficiently from one another while simultaneously allowing isolation of the effects under investigation. In particular, excessive adjustments within a treatment can lead to confounding effects. Therefore, factorial, intersubjective experimental designs were mainly used. Only individual design features were varied while everything else remained constant (e.g., male vs. female voice; Paper C). Second, a customized experimental platform for studying CI-mediated services interactions was developed to rule out potential confounding effects and to ensure high internal validity. The interface was kept neutral to exclude any confounding factors that could have led to undesirable effects on the key measures of interest (see Figure 4 for a comparative view of the text- and voice-based CI). In addition, demographic and personality constructs were collected in each experiment to rule out sampling-related influences. Finally, full randomization was applied to further control for sampling-related issues and for temporal effects. User panels from Prolific were used across the studies. The participants were routed from Prolific to an external survey tool, either Qualtrics or Unipark, which then randomly assigned the participants to an experimental treatment (see also Paper C, D, E & F).

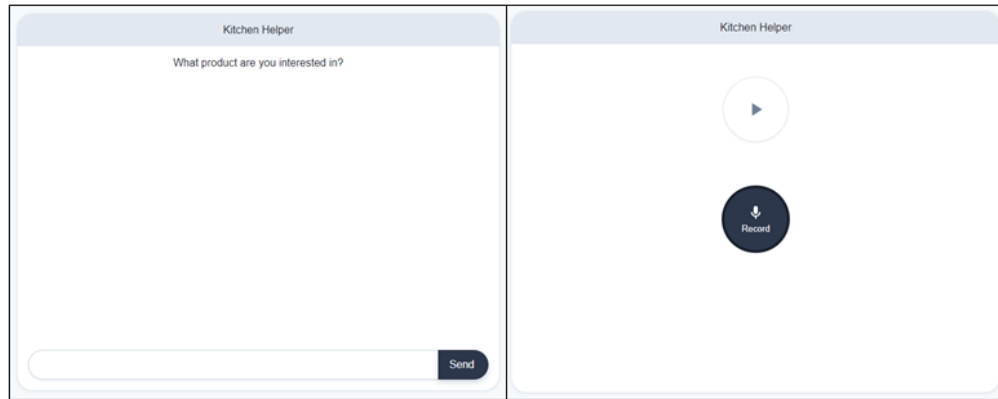


Figure 4. Text-based and voice-based version of experimental platform

Table 3 illustrates the dissertation's overall research design and summarizes some of the key design decisions taken. In summary, this dissertation (except RQ1, which takes an exploratory literature-based approach) builds on an explanatory-quantitative research strategy to analyze and provide causal evidence as to how CIs and their design affect user experience and service outcomes. That is, this dissertation mainly uses online experiments capturing cross-sectional data.

Table 3. Research design decision (adapted from Recker, 2013 and Cooper & Emory, 1991).

Spectrum	One end of continuum	Other end of continuum
Aim	Exploratory	Explanatory
Method	Qualitative	Quantitative
Boundary	Case	Statistical
Setting	Field	Laboratory
Timing	Cross-sectional	Longitudinal
Outcome	Descriptive	Causal
Ambition	Analyzing	Designing
<i>Decision on research design in this dissertation are highlighted in grey (especially with regard to the presented work on RQ2 and RQ3).</i>		

4 Results

This chapter presents the dissertation's major results. Section 4.1 provides a brief overview of the publications considered in this dissertation in light of the research questions posed. Section 4.2 provides a short summary of each paper's main findings.

4.1 Overview of Publications

Overall, the RQs posed in this dissertation are answered across six papers. Table 4 provides an overview of all publications that constitute this dissertation.

Table 4. Publications included in this dissertation

No.	Publication	Method	RQ
A	Zierau, N. ; Elshan, E.; Visini, C. & Janson, A. (2020) : A Review of the Empirical Literature on Conversational Agents and Future Research Directions. <i>In: International Conference on Information Systems (ICIS)</i>	Systematic literature review	1
B	Zierau, N. ; Engel, C.; Söllner, M. & Leimeister, J.M. (2020) : Trust in Smart Personal Assistants: A Systematic Literature Review and Development of a Research Agenda: <i>In: 15th International Conference on Wirtschaftsinformatik (WI 2020)</i> .	Systematic literature review	1
C	Zierau, N (2022): Gendered Interfaces as Surrogates of Trust? How Voice Gender and Communication Style Affect Interface Trust and Firm Perceptions. <i>Submitted to: Proceedings of the 56th Hawaii International Conference on System Sciences</i>	Experimental research	2
D	Tolmeijer, S.; Zierau, N. ; Janson, A.; Wahdatehagh, J. S.; Leimeister; J. M. & Bernstein, A. (2021) : Female by Default? – Exploring the Effect of Voice Assistant Gender and Pitch on Trait and Trust Attribution. <i>In: ACM CHI 2021 Late Breaking Work</i> .	Experimental research	2
E	Zierau, N. , Hildebrand, C., Bergner, A., Busquet, F., Schmitt, A. & Leimeister J.M. (2022) : Voice Bots at the Frontlines: Voice-Based Interface Enhance Interface	Experimental research	3

	Flow & Boost Service Outcomes. <i>Accepted for publication: Journal of the Academy of Marketing Science.</i>		
F	Schwende, M., Zierau, N. , Janson, A, Hammerschmidt, M. & Leimeister, J.M. (2022): “I Will Follow You!” – How Recommendation Modality Impacts Processing Fluency and Purchase Intention <i>Submitted to: International Conference on Information Systems (ICIS).</i>	Experimental research	3

This dissertation can be broadly divided into two parts. The first part is conceptual in nature and concerned with providing an integrative framework of CIs in service contexts (RQ1). To that end, Papers A and B each present the results of SLRs that were conducted to conceptualize both CI characteristics and respective service outcomes. Moreover, the empirical evidence for these interface characteristics on the service outcome were summarized and synthesized to provide a comprehensive agenda for research on user interactions with CIs.

The second part of this dissertation is empirically driven and addresses important characteristics of CIs pertaining both to the interface (i.e., voice gender; RQ2) and the interaction (i.e., modality; RQ3). Papers C and D present the results of a series of experimental studies that investigated the effect of voice gender on user experience and service outcomes, illuminating key boundary conditions in the process. Papers E and F present the results of a series of experimental studies that systematically assessed the effect of interaction modality (voice vs. text) on user experience and the downstream consequences of these effects for perceptual and behavioral service outcomes. Moreover, several boundary conditions are investigated, which are manifested in the important features of conversational design. Figure 5 provides an overview of how these papers contribute to the overarching conceptual framework of this dissertation (see Chapter 2).

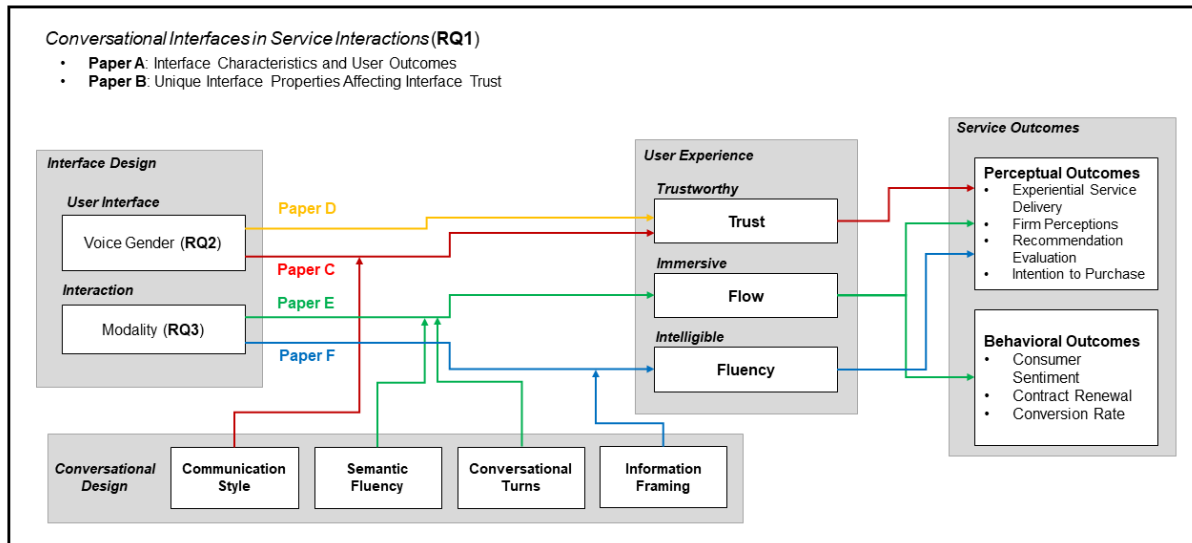


Figure 5. Overview of how papers contribute to the conceptual framework of this dissertation

4.2 Papers of the Dissertation

An overview of the research papers included in this dissertation is provided below. This includes bibliographical information and a brief explanation of the objectives, the research design, and the main results as well as each paper's respective contribution.

4.2.1 Paper A: A Review of the Empirical Literature on Conversational Agents and Future Research Directions

Citation

Zierau, N.; Elshan, E.; Visini, C. & Janson, A. (2020): A Review of the Empirical Literature on Conversational Agents and Future Research Directions. *In: International Conference on Information Systems (ICIS)*

Objective

Paper A presents the results of a SLR that was conducted to summarize and synthesize the empirical literature on CIs. As this review demonstrates, the literature on CIs has increased exponentially. Most publications have investigated the effects of selected interface characteristics or configurations on selected user perceptions and behaviors. Ultimately, this has led to a fragmented literature base. Specifically, the literature lacks an integrated conceptualization of CIs spanning technology and user-based perspectives, which makes it difficult to compare results and identify open questions. Hence, in this paper, an integrated framework was developed to conceptualize the characteristics (i.e., design features) of CIs and link them to categories of user experience outcomes.

Design and Results

To systemize interface characteristics and user outcomes and summarize the findings regarding the relationship between interface characteristics and user outcomes, a SLR was conducted. To provide a broad literature basis, 107 publications were identified based on a systematic outlet-based search in major journals and conference proceedings across relevant disciplines. These publications were analyzed using the abductive approach developed by Jeyaraj et al. (2006) that assumes a concept-centric perspective, according to which a conceptual framework was developed by iteratively coding independent variables, dependent variables, and their relationships. Consequently, five dimensions of interface characteristics (auditory, interaction, invisible, visual and verbal cues), five dimensions of user outcomes (rapport, social presence, trust, utility, and behavioral intention and outcomes), and the respective strengths of the relationships between the interface characteristics and user outcomes were identified. This analysis facilitated the development of a research agenda that is positioned in the research space that this framework creates.

Contribution to Dissertation

This paper conceptualizes and systematizes the key characteristics of CIs and links them to categories of user perceptions and behaviors. In doing so, the review creates an essential foundation for understanding service experiences with CIs. In the context of this dissertation, this publication has three main contributions. First, the review uncovers important categories of interface characteristics and user outcomes that form the service experience. Second, the review demonstrates that CIs and their design create specific user experiences with significant downstream consequences for users and firms. Third, the review highlights the critical psychological mechanisms that can explain these effects and uncovers important theoretical bases to consider.

4.2.2 Paper B: Trust in Smart Personal Assistants – A Systematic Literature Review and Development of a Research Agenda

Reference

Zierau, N.: Engel, C.; Söllner, M. & Leimeister, J.M. (2020): Trust in Smart Personal Assistants: A Systematic Literature Review and Development of a Research Agenda: *In: 15th International Conference on Wirtschaftsinformatik (WI 2020)*.

Objective

Paper B presents the results of a SLR that was conducted to summarize and synthesize the literature on trust in CIs. Anecdotal evidence indicates that users generally do not trust CIs (Kroque, 2017), citing the unique properties of these AI-based service interfaces that fundamentally affect the ways in which users engage with them (Pfeuffer, Benlian, et al., 2019). Among this paper's main objectives was to better understand which unique properties of CIs influence user trust. Trust is fundamental for the adoption and use of IS and allows far-reaching conclusions to be drawn regarding the user experience.

Design and Results

A SLR was conducted to better understand which factors in CIs influence user trust. To that end, 50 publications were identified using a database-based search strategy. The identified publications were analyzed in accordance with a concept-centric approach, and the results revealed three distinct perspectives based on the unique properties of CIs that drive trustworthy user experiences: user interface-driven trust, interaction-driven trust, and explanation-driven trust. Based on the framework developed in this paper, a comprehensive research agenda was developed to better understand how trust in CIs can be systematically built.

Contribution to Dissertation

This publication makes two essential contributions to this dissertation's overall objectives. First, based on the three distinct perspectives on trust in CIs developed in this paper, unique properties of these interfaces can be derived. Accordingly their anthropomorphic design, the dialogic interaction and their probabilistic nature shape user experiences with CIs. Second, it further highlights the critical role of specific design dimensions and their characteristics in creating trustworthy user experiences.

Overall, Papers A and B make three key contributions to RQ1. First, they elaborate on the fundamental properties of CIs from the users' perspective and the respective relevant

theoretical bases to consider. Second, these papers provide a holistic systematization of CIs' (design) characteristics and identify critical determinants of service outcomes. Third, the strengths of the empirical evidence regarding these characteristics' influence on user outcomes are outlined and respective gaps are highlighted.

4.2.3 Paper C: Gendered Bots as Surrogates of Trust? How Voice Gender and Conversational Style Affect Interface Trust and Firm Perceptions

Zierau, N (2022): Gendered Bots as Surrogates of Trust: How Voice Gender and Conversational Style Together Affect Interface Trust and Firm Perceptions. *Submitted to: Proceedings of the 56th Hawaii International Conference on System Sciences*

Research Objective

Paper C presents the results of an experimental study designed to investigate the combined effects of voice gender and communication style on interface trust and firm perceptions. Previous literature has shown that the influence of voice gender on interface trust is highly context-dependent (Nass et al., 1997). This paper addresses two specific research gaps. First, it investigates the effects of a novel boundary condition manifested in the conversational design of voice-based interfaces (i.e., the conversational style an interface employs). Second, it examines the downstream consequences of interface trust on the firm the voice-based CI represents.

Design and Results

A 2 (voice gender: female vs. male) x 2 (communication style: social-oriented vs. task-oriented) between-subject online experiment was designed to investigate the research gaps. Specifically, the participants were asked to file an insurance claim using the respective interface conditions. The results show that trust is lower in male voices when using a social-oriented conversational style. However, for a task-oriented conversational style, the gender of the voice has no significant effect on interface trust. Moreover, the results indicate that reduced levels of interface trust experienced when interacting with a voice-based CI using a male voice and a social-oriented conversational style results in a more negatively-valenced service experience and less attributions of benevolence towards the firm deploying the interface.

Contribution to the Dissertation

These results provide a nuanced perspective on how voice gender in CIs affects user experience and service outcomes. This paper extends previous research showing that imbuing interfaces with a voice can affect trust attribution following social heuristics (Nass et al., 1997). Specifically, this publication makes several significant contributions to this dissertation's overall objectives. First, this paper reveals a critical boundary condition (i.e., communication style) that influences the effect of voice gender on trustworthy user experiences. Second, the results demonstrate that a more trustworthy interface

experience when using a voice-based CI translates into more beneficial service outcomes for firms (i.e., service experience, firm benevolence). These findings provide important insights into the theory-driven voice design of CIs.

4.2.4 Paper D: Female by Default? – Exploring the Effect of Voice Assistant Gender and Pitch on Trait and Trust Attribution

Citation

Tolmeijer, S.; Zierau, N.; Janson, A.; Wahdatehagh, J. S.; Leimeister, J. M. & Bernstein, A. (2021): Female by Default? – Exploring the Effect of Voice Assistant Gender and Pitch on Trait and Trust Attribution. In: *ACM CHI 2021 Late Breaking Work*.

Objective

Paper D's main objective was to investigate the effect of non-gendered vs. gendered voices on stereotypical traits and trust attribution across different task contexts. This study builds directly on a limitation of prior research (including Paper C), that is, to conceptualize voice gender only in a binary fashion. Specifically, technology providers have started to develop gender-ambiguous voices to provide more inclusive user experiences (Q, 2019; Sutton, 2020). Moreover, gender-ambiguous voices are supposed to stop the reinforcement of harmful gender stereotypes (UNESCO, 2019). This paper builds on Paper C and provides further evidence on the effects of gendered voices in CIs on user experience and service outcomes.

Design and Results

To investigate the effect of gendered vs. non-gendered voices on user perceptions, a 5x2 (voice type x task type) factorial web experiment was designed. Both high-pitched and low-pitched female and male voices were used for the voice type. Moreover, the study uses a gender-ambiguous voice that could not be easily categorized as female or male. The task type was also varied across voice conditions between an assistance task (i.e., booking a flight) and a compliance task (i.e., responding to sensitive questions). These tasks were selected according to traditional gender stereotypes, according to which men are regarded as leaders and women as assistants. The results revealed no direct effects of voice type on stereotypical trait attribution and trust. However, the results reveal that task context was instrumental in stereotype activation. Interfaces using male voices were perceived as stereotypically male in the compliance task, while interfaces using female voices received significantly more trust in the assistance task.

Contribution to Dissertation

The results of this paper add to Paper C in two important ways. First, this paper further highlights that the effect of voice gender is highly context-dependent. Specifically, it shows that the type of task can elicit stereotype-informed trust attributions. Second, this

paper did go beyond a binary gender conceptualization by evaluating user perceptions of a gender-ambiguous voice. The results indicate that there are no systematic differences in trust attribution between gendered and non-gendered voices in CIs, providing initial evidence that gender-ambiguous voices represent an interesting alternative to gendered voices in service contexts.

Overall, Papers C and D provide a comprehensive look at the subtle effect of voice gender and make several key contributions to RQ2. First, evidence for a novel boundary manifested in the conversational design is provided (i.e., conversational style). Second, the downstream impact of voice gender on firm perceptions is investigated. Finally, the effect of gender-ambiguous vs. non-gendered voices in CI is systematically examined.

4.2.5 Paper E: Voice Bots at the Frontlines – Voice-Based Interface Enhance Interface Flow & Boost Service Outcomes

Citation

Zierau, N., Hildebrand, C., Bergner, A., Busquet, F., Schmitt, A. & Leimeister J.M. (2022): Voice Bots at the Frontlines: Voice-Based Interfaces Enhance Interface Flow & Boost Service Outcomes. *Accepted for publication: Journal of the Academy of Marketing Science*.

Objective

Publication E's main objective was to systematically investigate the effect of communication modality (voice vs. text) on interactivity-related affective user experiences and reveal the downstream consequences for users and firms, thus contributing to RQ3. In recent years, service firms have increasingly used voice-based CI to automate selected service encounters (Diederich et al., 2022; Seaborn et al., 2021). This burgeoning trend raises the question of how this novel interaction paradigm affects users in service contexts, how users perceive firms using these interfaces, which psychological mechanisms might account for these effects, and what the downstream consequences for users and firms are.

Design & Results

To that end, a novel conceptualization of voice-based CI in service contexts was developed based on a review of prevalent communication theories (i.e., media richness). Across four experimental studies, it became clear that the affordances of voice-based CIs (i.e., verbal cues, channel synchronicity) increase interface flow and in turn lead to a more positively valanced service experience and boost service firm outcomes (i.e., contract renewal, conversion rate, and consumer sentiment) during the process of filing insurance claims. Moreover, two important boundary conditions were identified as having manifested in fundamental aspects of conversational design (i.e., semantic disfluency and the number of conversational turns). Specifically, these results demonstrate that the effect of voice-based CI s on interface flow is reinforced when semantic fluency is high and when the number of conversational turns is low.

Contribution to Dissertation.

This paper provides substantial empirical evidence for the effect of voice-based (as opposed to text-based) CIs on user perceptions and service outcomes. This publication makes several critical contributions to this dissertation's overall objectives. First, the

results demonstrate that voice-based CIs can lead to more immersive user experiences in customer service settings, translating into beneficial service outcomes for firms. Second, this paper reveals the important boundary conditions manifested in communication modality that influence the effect of voice-based CIs on user experience.

4.2.6 Paper F: “I Will Follow You!” – How Recommendation Modality Impacts Processing Fluency and Purchase Intention

Citation

Schwede, M., **Zierau, N.**, Janson, A, Hammerschmidt, M. & Leimeister, J.M. (2022): “I Will Follow You!” – How Recommendation Modality Impacts Processing Fluency and Purchase Intention. *Submitted to: International Conference on Information Systems (ICIS)*.

Objective

In recent years, voice-based (as opposed to text-based) CIs have been used increasingly to promote and sell products and services (Tricks, 2021). However, research is lacking with respect to how different recommendation modalities impact user experiences and related service outcomes and how recommendations should be designed across modalities. As such, Paper F’s key objectives were to systematically investigate the effect of communication modality (voice vs. text) on interactivity-related cognitive user experiences, to reveal the downstream consequences for users and firms, and to identify important boundary conditions to these effects.

Design and Results

To investigate the impact of recommendation modality and design on interface experiences (i.e., processing fluency) and service outcomes (purchasing intention), two web experiments were designed. In the first study, people were tasked with receiving a food recommendation either using a text-based or voice-based CI. In the second study, using the same experimental paradigm, the interaction was manipulated to include either no explanation, a verbal-framed explanation, or a numerical-framed explanation across the two modalities, resulting in a 3×2 factorial design. These two studies demonstrated that recommendations provided by text-based interfaces are more easily processed than those provided by voice-based CIs. Moreover, the results indicate that higher processing fluency reduces risk perceptions and increases recommendation persuasiveness, ultimately leading to higher purchasing intentions. Finally, this paper’s results demonstrate that numerically framed recommendations enhance fluency perceptions as distinct from verbally framed recommendations or no recommendations at all.

Contribution to Dissertation.

This paper extends Paper E's findings in several key ways. First, this paper shows that the interactivity-related affordances of voice-based CIs can be also associated with disadvantages for users. Specifically, this paper demonstrates that processing fluency is lower in voice-based CIs, ultimately reducing purchasing intention. Second, this paper examines another critical aspect of conversational design—information framing (numerical vs. framing). Third, this study investigates the impact of voice-based interfaces in a novel service context (i.e., receiving information instead of providing information, as in Paper E).

Overall, Papers C and D provide a comprehensive look at the effect of interaction modality and make several key contributions to RQ3. First, these papers investigate the impact of voice-based CIs on important perceptual and behavioral service outcomes. Second, the differential effect of two critical psychological mechanisms (i.e., flow, processing fluency) that are based on the unique affordance of voice-based CIs are uncovered. Finally, these papers provide evidence for several boundary conditions representing important features of conversational design (i.e., semantic fluency, number of conversational turns, information framing).

5 Discussion and Outlook

This section highlights this dissertation's core contributions, identifies the implications for research and practice, elaborates on the limitations of this research, and discusses future research avenues.

5.1 Summary of Results

This dissertation aims to advance our knowledge on the emergent phenomenon of the use of CIs in service contexts. Specifically, this work contributes to our understanding of how these interfaces and their design affect user experiences and service outcomes in three major ways:

First, Paper A and B together provide a holistic conceptualization of CIs in service contexts. Accordingly, the unique properties of CIs (e.g., human-likeness, interactivity, intelligence) shape user experiences and inform service outcomes. Specifically, the features of CIs may be distinguished into five dimensions (verbal, visual, auditory, interaction, and invisible cues) that together form the user experience. The strength of the empirical evidence for different dimensions of service outcomes (relationship, trust, social presence, behavioral intentions, and outcomes) was also highlighted. In this process, a comprehensive agenda for research on CIs in the service context was developed from a user-centered perspective. Thus, these two papers form a foundation on which the other components of the dissertation may build (see Paper A & B).

Second, Paper C and D adopt a user interface-focused perspective and provide an in-depth examination of voice gender in CIs. Paper C introduces a novel boundary condition showing that male voices are trusted less if paired with a social-oriented communication style (as opposed to a task-oriented communication style). These differences in interface trust ultimately led to a more negatively-valenced service experience and less attributions of benevolence toward the firm. No systematic differences between voice gender and interface trust can be observed when using a transactional voice. Paper D shows that trust and stereotypical trait attribution based on voice gender depends on the task context. Female voices were trusted more in an assistance task, while male voices were ascribed male attributes in a compliance task. Moreover, this paper hints at the potential for using gender-ambiguous voices in service contexts to mitigate gender stereotyping. No significant differences were observed in user perceptions between non-gendered and gendered voices (see Paper C & D).

Finally, Paper E and F together assume an interaction-focused perspective and systematically investigate the role of interface modality and communication design on user experiences and service outcomes. Paper E demonstrates that the unique properties of voice-based (as opposed to text-based) CIs (i.e., verbal cues, channel synchronicity) increase interface flow during the process of filing an insurance claim, resulting in a more positive and positively-valenced service experience affecting a wide range of service outcomes (i.e., consumer sentiment, contract renewal, and conversion rates). Moreover, this paper also provides two important boundary conditions that manifested in conversational design and reduced flow in voice-based interfaces (i.e., semantic disfluency and the amount of conversational turns). Adopting a complementary perspective, Paper F provides evidence that the same properties diminish processing fluency when receiving a product recommendation, with negative downstream consequences for recommendation evaluation (risk perceptions, recommendation persuasiveness) and intention to purchase. Moreover, this paper demonstrates how recommendation modality interacts with the framing of recommendation explanations (verbal vs. numerical), indicating that numerical recommendations are a potent measure that may be implemented to increase fluency and intention to purchase across conditions (see Paper E & F).

Taken together, these results provide support for the overarching assumptions raised in the conceptual framework of this dissertation on how CIs and their design affect user experiences and service outcomes (see Figure 3). First, they indicate that the characteristics of CIs (e.g., their anthropomorphic appearance, dialogic interactivity) uniquely influence user experiences. Specifically, it is shown that voice gender (in interplay with communication style) affects interface trustworthiness and that interaction modality can affect both the immersion (e.g., flow) and intelligibility of the interaction (e.g., processing fluency). Second, these results show that the unique properties of conversational interfaces impact how CI designs affect user experiences. For instance, it is shown that conversational style moderates the effect of voice gender on interface trust (see Paper C) and that the number of conversational turns moderates the effect of interface modality on flow-like user experiences (see Paper E). Finally, these findings provide evidence that the user experience is instrumental for shaping downstream effects on both perceptual and behavioral service outcomes (see Paper C, E, & F).

A summary of the results across all publication is provided in Table 5.

Table 5. Summary of results across all papers.

RQ	Paper	User Experience	Service Outcomes
1	A	CI characteristics can be distinguished into five dimensions (verbal, visual, auditory, interaction, and invisible cues).	- Service outcomes of CIs can be distinguished into rapport, trust, social presence, behavioral intentions, and outcomes. - The strength of the empirical evidence of interface characteristics on these outcomes is highlighted.
	B	There are three unique properties of CIs (i.e., human-likeness, dialogic interactivity, and intelligence) that shape (trustworthy) user experiences.	
2	C	- For a social-oriented communication style, trust in male voices is decreased - For a task-oriented communication style, there are no systematic differences based on voice gender.	Enhanced levels of trust experienced when interacting with a voice-based CI using a female voice and a social-oriented style lead to a more positive overall service experience and benevolence attributions toward the firm.
	D	- For an assistance task, trust in female voices is enhanced. - For a compliance task, male voices are assigned more stereotypical male traits. - There are no systematic differences between a gender-ambiguous and gendered voices concerning interface trust.	
3	E	- Voice-based (as opposed) to text-based CIs enhance interface flow. - The amount of conversational turns decrease flow in voice -based CIs. - Semantic disfluency decreases flow in voice-based CIs.	- Enhanced levels of flow in voice-based CIs lead to a more experiential delivery experience and a more positive firm evaluation. - Enhanced levels of flow in voice-based CIs lead to more positive behavioral service outcomes (i.e., consumer sentiment, contract renewals, and conversion rates).

	F	• Voice-based (as opposed to text-based) CIs decrease processing fluency. • Numerical (as opposed to verbal) framed information enhances processing fluency.	• Decreased levels of processing fluency when using a voice-based CI lead to a more negative recommendation evaluation and decreased levels of intention to purchase.
--	----------	---	---

5.2 Theoretical Implications

This dissertation's findings make several novel contributions to emerging work on CIs' impact on user experiences and service outcomes and the impact of AI-driven interfaces on the service value chain as well as novel insights into the psychological mechanisms of CIs on the technology-augmented user in service contexts. In the section that follows, these contributions are discussed along with the three research questions posed at the beginning of the dissertation.

5.2.1 RQ1: Toward a Holistic Conceptualization of Conversational Interfaces in Service Contexts

With regard to RQ1, this dissertation (i.e., specifically Papers A and B) makes two novel contributions. First, the results provide a unified conceptualization of CIs in service contexts. Accordingly, three high-level insights into CIs may be drawn based on this conceptualization. First, building on and extending existing conceptual work on AI-based technologies (Maedche et al., 2019), this dissertation shows that CIs represent a novel form of service interface featuring unique properties (human-likeness, interactivity, ambiguity). Second, this dissertation provides a fresh perspective on existing work regarding the systematization of CI features (Feine et al., 2019). Finally, a systematization of service outcomes for CI-mediated encounters is provided and the empirical evidence for how CIs and their design impact these outcomes is illuminated. Overall, these results provide an important foundation for understanding user experiences with CIs in service contexts, fostering the conceptual convergence of a multi-faceted and complex research phenomenon (see Paper A and B).

Second, this dissertation's results provide an integrated summary of the progress of research on CIs in service contexts across different research disciplines. Based on the analysis of the selected studies, this dissertation proposes various research avenues. Taken together, these research agendas serve as a comprehensive guide for researchers working on IS, marketing, and HCI to advance research on CIs in service contexts. At the same time, the coded literature can support practitioners in implementing empirically

substantiated design decisions (Diederich et al., 2022). For instance, service designers can draw on the literature basis presented in Papers A and B to inform user interface designs with the aim of increasing user trust (Nunamaker et al., 2011) or to identify conversational strategies with the aim of improving the interaction flow (L. Clark et al., 2019).

5.2.2 RQ2: The Effects of Voice Gender on User Experiences and Service Outcomes

Concerning RQ2, this dissertation (specifically Papers C and D) makes three novel contributions. First, these studies build on previous work showing that the effect of voice gender on interface trust is highly context-dependent (Nass et al., 1997; Tay et al., 2014). Specifically, this dissertation illuminates a novel boundary condition manifested in the specific characteristics of conversational design (i.e., conversational style). The results indicate that users evaluate the effect of voice gender on trust differently based on the interface's specific words. These results provide a new look at the component fit theory, showing that the features of the voices interact with the spoken to create a unique user experience (Isbister & Nass, 2000). Moreover, this dissertation confirms previous work showing that stereotype informed voice gender- task congruence increases trust in the CI (Nass et al., 1997). Specifically, this dissertation distinguishes between a compliance task (stereotypical male) and an assistance task (stereotypical female) (see Paper C and D).

Second, the present study's findings provide novel insights into user–technology relationships (Hoffman & Novak, 2018; Novak & Hoffman, 2018; Schuetz & Venkatesh, 2020) and how gendered interfaces create new “assemblages” of user experiences. Specifically, this research provides evidence that gender cues in voice-based CIs trigger the social heuristics known from interpersonal interactions. These judgments not only impact the interface itself (i.e., the relationship between user and interface) but also the firm providing the interface experience. As such, this dissertation provides evidence that anthropomorphically designed CIs can create spill-over effects toward service firms (see Paper C).

Finally, this dissertation is the first to design and empirically evaluate a gender-ambiguous voice. The results (i.e., Paper D) indicate that gender-ambiguous voices are on a par with gendered interfaces with respect to user experiences, hinting at their potential for service provision in terms of mitigating the potential adverse effects of gendering. At the same time, this dissertation also shows that most people have no mental frame

for a third gender (at least based on voice). Although the pitch was in a range that impeded the clear identification of gender, the agent was nonetheless classified as either male or female. This raises the question of whether gender-ambiguous voices can be successfully designed at all and indicates the need for further research on the extent to which gender-ambiguous voices can actually prevent gender stereotyping (see also Section 5.5) (see Paper D).

5.2.3 RQ3: The Effects of Interaction Modality on User Experiences and Service Outcomes

With regard to RQ3, this dissertation (specifically Papers E and F) makes three new contributions. First, the present work is one of the first to empirically contrast the effects of a voice-based CI with those of a text-based CI. On that basis, this dissertation develops a novel theory of voice-mediated service interactions, demonstrating that the affordances of voice-based CIs provide a richer and more synchronous user experience. This dissertation also provides evidence on the important role that interface experience plays as an underlying mechanism to explain these effects. Specifically, we provide a fresh perspective on existing work regarding flow (Agarwal & Karahanna, 2000; Ghani et al., 1991) and processing fluency (H. Song & Schwarz, 2008; Thompson & Ince, 2013), highlighting that interaction modality is a key driver of absorbing and intelligible user experiences. To that end, Paper E shows that voice-based CIs increase flow-like user experiences, which in turn create more experiential service delivery experiences and positively affect service outcomes. However, Paper F demonstrates that voice-based CIs reduce processing fluency, which in turn diminishes recommendation outcomes and negatively affects purchase intention (see Paper E and F).

Second, the present work introduces and empirically tests three novel boundary conditions manifested in the “conversational design” of service interactions. First, this work demonstrates that using easy-to-process language is instrumental in enhancing interface flow in voice-based CIs with positive downstream consequences for both perceptual and behavioral service outcomes. Second, this work also demonstrates that a higher number of conversational turns can negatively impact interface flow in voice-based CIs, as they cause the communication to become more fragmented. Third, the results of this dissertation indicate that the provision of numerically framed explanations is critical to boosting processing fluency and, in turn, positively affects recommendation outcomes (persuasiveness, risk) as well as purchasing intention. Together, these results provide novel insights into the conditions under which voice-based CIs can positively or negatively

affect user experiences. Moreover, these findings highlight the critical role of conversational design in creating satisfactory user experiences and boosting service outcomes (see Paper E and F).

Third, the present study's findings also provide a novel perspective on user–technology relationships (Hoffman & Novak, 2018; Novak & Hoffman, 2018; Schuetz & Venkatesh, 2020) and the underexplored role of interface modality in shaping user experiences and service outcomes. Specifically, these results demonstrate that natural interaction through a voice-based CI causes both positive attributions (such as enhanced interface flow) and negative attributions toward the user–technology “assemblage”. In spite of its ability to extend the capacities of traditional service interfaces through the voice modality, the formation of the assemblage between the user and the service interface diminishes processing fluency. Thus, the capacity of voice-based CIs reported herein illustrates how interaction modality as a key conceptual feature of CIs evoke both nuanced positive (i.e., greater flow) and negative effects (i.e., lower processing fluency) (see Paper E & F).

5.3 Managerial Implications

This dissertation's findings have immediate implications for service designers and firms. The section that follows provides an overview of these implications and design guidelines for industry practice. These implications are categorized into insights targeted at the general deployment of interfaces across the service channel and those that support the efficient design of the user interface or the conversational interaction. Most of these implications are mirrored from the publications in part B of this dissertation.

5.3.1 Toward an Effective Deployment of Conversational Interfaces

This dissertation's results provide insights into the effective deployment of CIs across the service value chain as follows:

Conversational interfaces facilitate a more satisfactory and inclusive user experience. This dissertation shows that CIs can create a more satisfactory user experience. According to earlier research, traditional interfaces can lead to a sense that the human touch is lacking (van Doorn et al., 2017). However, the use of judiciously selected CIs can create a more sensory-rich user experience across different tasks and contexts. Moreover, voice-based interfaces can be used to provide access disadvantaged user groups (for example, users who are blind or visually impaired; Abdolrahmani et al., 2018). Thus,

CIs can be used not only to provide a more sensory-rich and human-like user experience but also to enhance inclusivity across the service value chain (Paper E).

Choose interface type according to the requirements of the task. In choosing the appropriate self-service channel, firms are often driven by exploratory behavior rather than systematically linking the service task's requirements to the affordances of a given interface (modality). This dissertation offers guidelines for the effective selection of the appropriate service channel. Specifically, the present research offers guidance regarding the selection of voice-based (as opposed to text-based) CIs (e.g., see Paper E for a summary of the properties of voice-based as opposed to text-based CIs). Voice-based CIs can be effectively deployed for those tasks that benefit from high channel synchronicity and a rich symbol set, such as the provision of claim-related information. By contrast, tasks that require users to deliberate are better suited to text-based exchanges. Specifically, service tasks that are associated with high risks or high cognitive demands, such as the selection of a complex, high-priced product, are better suited to asynchronous communication (i.e., text-based communication) (Paper E).

Conversational interfaces can be efficiently integrated into firms' existing service operations. The rapid development of technology platforms for designing CIs, such as Google Dialogflow or Amazon Alexa, allows users with no or few programming skills not only to easily develop CIs but also to integrate these into the current IT infrastructure and rapidly deploy them across service channels. Service firms can easily adapt existing interfaces and build on prevalent language models instead of building firm-owned native CIs. As such, the development, technical integration, and deployment are *effective* (i.e., the infrastructure is provided, maintained, and continuously improved by providers, such as Amazon or Google), *customizable* (e.g., employees can easily adapt the interface according to the firm's preferences in a low-code environment), and *adaptive* (e.g., firms can design and deploy different interface designs following rapid development cycles). In short, the use of CIs is highly modular, and they can be easily integrated into different service encounters along the service value chain (Paper E).

5.3.2 Toward an Effective Design of Voice

The results presented herein provide insights into the effective design of CIs. Specifically, the results demonstrate that the design of (gendered) voices can be critical to enhancing both user experience and service outcomes. The section that follows discusses three principles of "voice design":

Adapt voice (features) to context. The findings of this dissertation show that the effect of voice gender is highly context-dependent. For instance, Paper C demonstrates that the interface's conversational style affects how users perceive voice gender. These findings illustrate the importance of a contextual perspective on CI design. Specifically, service designers need to carefully orchestrate voice features and characteristics of conversational design, considering task characteristics and user preferences. In this regard, creating and evaluating personas with users could help create compelling voice personalities (Braun et al., 2019) (see also Paper C).

Personalize voices to user preferences. The results of this dissertation show that user characteristics (e.g., user gender) can affect the perception of voice gender of CIs (e.g., women trust the gender-ambiguous voice more than men, see Paper D). Hence, service designers should carefully test and ensure that the voice of the interface is tailored to the preferences of the user. It can also be useful to allow users to choose between different voices. Many CIs, such as Apple's Siri, provide users with the option to choose different voices. Based on user agency, a positive effect of choice with respect to different voices on user experience and service outcomes is to be expected.

Experiment with gender-ambiguous voices. This research contributes to the discourse on gendered AI by providing further evidence that the positive effects of voice gender on the user experience is highly context-dependent (and in some contexts non-existent). Thus, this dissertation's findings call into question the prevalent industry practice of always choosing female over male voices. Moreover, the findings indicate that gender-ambiguous voices are on a par with gendered voices. As the use of female voices is linked to the reinforcement of harmful gender stereotypes (UNESCO, 2019), these findings should provide service firms with further impetus to imbue their voice assistants with more diverse demographic features, reflecting the diversity of society, or to use gender-ambiguous voices (Paper D).

5.3.3 Toward an Effective Design of Conversational Interactions

This dissertation's results indicate that the semantic and structural characteristics of natural language-based interactions are critical for boosting user experience and service outcomes. Most importantly, the results indicate that text-based and voice-based interactions should follow a different logic of conversation. The section that follows discusses four principles of "conversational design".

Optimize for semantic fluency in voice-based interfaces. This dissertation's findings indicate that semantic fluency differentially affects interface experiences in voice-based

than text-based CIs. For voice-based interfaces, the results indicate that service designers should reduce semantic fluency by using more familiar “words”, simplifying syntactic structures, and adopting a less formal conversational style. Conversely, the results indicate that text-based exchanges are less sensitive to semantic fluency. As a result, service designers can even resort to more complex formulations in text-based exchanges if needed but should carefully optimize semantic fluency in voice-based exchanges (Paper E).

Use few conversational turns in voice-based exchanges and many in text-based exchanges. This dissertation has demonstrated that user satisfaction when using CIs depends on the number of conversational turns into which a service interaction is structured. Specifically, service designers should limit the number of conversational turns and carefully avoid disjointed interaction when designing voice-based interfaces. By contrast, when designing for text-based exchanges, division of the overall service encounter into more rather than fewer conversational turns enhances flow-like experiences and positively affects service outcomes (Paper E).

Use numerical framing of information. The results of this dissertation indicate that users more easily process information that is presented in a numerical as opposed to a verbal fashion. These effects are constant across modalities. These highlights the importance of presenting information in a concrete manner when using CIs. In particular, service designers should provide numerically framed information to support users in easily processing messages (Paper F).

Adapt conversational style to context. This dissertation’s findings indicate that a social-oriented communication style can trigger stereotypical trait attributions toward CIs. Especially in contexts where attributing certain personality traits to a CI is beneficial to the user experience (e.g., in educational contexts), it could be helpful to use social-oriented languages. However, for sensitive contexts, companies should choose a task-oriented communication style to avoid the risk of serving harmful stereotypes and to create a more inclusive user experience (Paper C).

Overall, these theory-driven principles suggest that the conversational design of CIs is sensitive to the use of semantics, the number of conversational turns, the framing of information and the communication style used. This indicates that service designers should carefully test the design of conversational service interactions against the requirements of the user and the task to optimized user experience and boost service outcomes.

5.4 Limitations

The research conducted in this dissertation has several limitations relating to the underlying data, the research methodology, and the general validity of the results. In this section, only those limitations that are associated with the general research strategy and approach are discussed. The individual papers provide further detail on the limitations of the work presented herein.

First, the dissertation's findings rely heavily on web experiments involving participants engaging in an artificial scenario. As such, this dissertation can provide high internal validity (Recker, 2013). However, although the scenarios were designed to be as realistic as possible, it is not clear whether the results will be realized in the same way when CIs are deployed outside the lab. Although initial studies have shown that text-based CIs also positively influence service outcomes outside the laboratory (Schanke et al., 2021), few studies have investigated the influence of voice-based CIs on users and companies on the basis of field data. Hence, future researchers should aim to replicate and extend the dissertation's findings by conducting field experiments or analyzing field data.

Second, this dissertation has mainly followed a behavioral approach by generating several insights that reveal the causal effects of CIs on user perceptions and service outcomes. However, only tentative claims may be made regarding the utility of CIs and their design in relevant application contexts. Hence, future research following a design science paradigm (Gregor & Hevner, 2013) could focus on creating application knowledge for the successful use of CIs along the service value chain. For example, design principles could be systematically derived for the trustworthy design of CIs in highly sensitive contexts (e.g., eliciting personal health-related information) or for the conversational design of voice-based CIs in various service settings.

Third, while users increasingly encounter CIs across the service value chain, this technology remains relatively new and immature. This may affect this dissertation's results in two key ways. First, some of the user perceptions especially regarding the voice-based CIs may be driven by novelty effects. Despite controlling for experience with these interfaces, the relative novelty of this technology may have influenced user perceptions and behaviors (Tsay et al., 2020). As users become increasingly experienced in using voice-based technologies, people may differently evaluate the user experience and adjust their behavior. Second, since the technology is still developing, the limited range of application may also have influenced the results (i.e., in the experiments conducted in

this dissertation users went through a pre-defined interaction). Therefore, as the technology becomes more ubiquitous and its technological basis reaches maturity, this dissertation's findings should be continually reviewed in novel contexts.

5.5 Future Work

CIs are here to stay and are only expected to grow in terms of their importance for service provision (Huang & Rust, 2021). In this regard, the present dissertation opens four new and significant research avenues on the ongoing infusion of CIs into service encounters. For each avenue, this section motivates and formulates specific directions that will guide future research activities to address the identified gaps. Some of these research directions are mirrored from the publications in part B of this dissertation. Table 6 summarizes these avenues and offers directions for further research that will advance our understanding of conversational interfaces in service contexts:

Table 6. Agenda for future research on conversational interfaces in service contexts

Research Avenue	Research Directions
Avenue 1: Conversational Interactions	• Direction 1.1: Explore when and how interaction modality (voice vs. text) affects user input and output. • Direction 1.2: Investigate the impact of device type in conversational interactions on user experiences and service outcomes. • Direction 1.3: Investigate how conversational interfaces and their design affect user perceptions and service outcomes over time.
Avenue 2: Designing Conversational Service Interactions	• Direction 2.1: Develop a method to systematically design voices for conversational interfaces. • Direction 2.2: Identify and study successful configurations of conversational user interfaces. • Direction 2.3: Develop methods to systematically design conversational service interactions. • Direction 2.4: Investigate designs of human-in-the-loop approaches for automated conversational service encounters.
Avenue 3: Potential of Voice-Based Service Channels	• Direction 3.1: Investigate how voice analytics can be successfully deployed across the service value chain. • Direction 3.2: Explore for which service tasks a voice interface is a suitable addition to a multi-channel offering. • Direction 3.3: Explore when and how the sharing of vocal (as opposed to textual) data affects user experiences and service outcomes.
Avenue 4: Ethical Implications of Gendered Interfaces	• Direction 4.1: Analyze when and how anthropomorphized interfaces affect users' mental models and the adverse effects on users' attitudes and behaviors. • Direction 4.2: Assess the unintended downstream consequences of gendered interfaces for interpersonal interactions and identify ways to mitigate these.

5.5.1 Avenue 1: Conversational Interactions

As this dissertation has demonstrated, CIs fundamentally affect how users interact with IS. However, several gaps remain with respect to the effect of CIs on user experience. Considering the increasing importance of CIs and the lack of research on how these interfaces affect user experience, the following four research directions are proposed:

Direction 1.1: *Explore when and how interaction modality (voice vs. text) affects user input and output.* In the scope of this dissertation, only the user outcomes relating to interaction modality that are most relevant to service firms were investigated (i.e., contract renewal, consumer sentiment). However, multiple user outcomes that are proximal to the interface experience have not been considered in this dissertation. According to theory regarding the unique affordances of voice- vs. text-based communication (see Paper E), it is likely that the use of voice-based CIs fundamentally affects the ways in which users both provide and process information across tasks and contexts (J. A. Berger et al., 2021; King et al., 2021). Specifically, future researchers should investigate when and how voice-based modalities affect information disclosure to machines (e.g., sensitive information). Another fruitful research avenue could be to assess whether particular information items (i.e., in recommendations or news) are perceived differently in voice-based vs. text-based interfaces CIs.

Direction 1.2: *Investigate the impact of device type in conversational interactions on user experiences and service outcomes.* Existing research has been mostly agnostic regarding the type of device (e.g., smartphone vs. personal computer) used to conduct automated service interactions (see Melumad & Meyer, 2020), notably demonstrating that smartphones (vs. personal computers) elicit greater disclosure owing to greater perceptions of psychological comfort. However, different device types offer distinct affordances that may affect how the same service interaction is experienced. In this context, it is particularly important to consider the differences between personal computers and smartphones as the currently dominant devices (Melumad & Meyer, 2020). Smartphones differ from personal computers not only in that they have a much smaller keyboard (which affects the speed at which a message can be written), but they tend to be used in closer proximity. Furthermore, smartphones can be used anywhere, even in public spaces, while users must sit down to perform service tasks using a personal computer (Melumad & Meyer, 2020). In this context, two exciting research avenues emerge. First, the differences between smart phones and personal computers could significantly influence user preferences regarding interaction modality (i.e., voice vs. text). Second,

the device could also influence users' expectations and requirements for interaction design. Smartphones, for example, could induce expectations of interaction that is as human-like as possible.

Direction 1.3: *Study how conversational interfaces and their design affect user perceptions and service outcomes over time.* In this dissertation, selected user perceptions and behavior were studied at a discrete point in time. Hence, the present study does not provide any insights on the evolutionary nature of user–technology relationships (McKnight et al., 2020). Specifically, the scope of this research did not encompass the temporal effects of CIs on user experience and service outcomes. For instance, while some design features may drive user perception and behavior in the initial adoption phase, others may become more important in the post-adoption stage. Moreover, some user experience dimensions, such as user trust, require multiple interactions for reliable formation (McKnight et al., 2020; Söllner & Pavlou, 2016). Hence, future research could employ a longitudinal perspective to study the effects of CIs and their design on user experience and service outcomes. Panel studies could provide opportunities to study changes in relevant user experiences at different points in the user journey. Simulation-based approaches may also be useful in providing a processual perspective (Paper A).

5.5.2 Avenue 2: Designing Conversational Service Interactions

This dissertation has demonstrated that designing satisfying service encounters with CIs can be a complex endeavor. Many independent but interrelated design aspects of CIs must be considered. However, theory-driven and empirically validated guidelines on CI design remain scarce. In this regard, four research directions emerge:

Direction 2.1: *Develop a method to systematically design voices for conversational interfaces.* Another promising future research avenue lies in the effective “voice design” of CIs. This dissertation provides evidence that gender-specific voice features drive user experience with important downstream consequences for users and firms. Specifically, voices are becoming focal points for the formation of user experience (see Paper C). To that end, congruity between the trait associations that a certain voice and its features evoke in a user and the (intended) character and firm personality is becoming an important design goal. As such, many technology providers are providing the option to creating customized voices (Schmitt, Zierau, et al., 2021; Sutton et al., 2019). Hence, future research could systematically assess how to map a voice and its features to a brand's personality or character (Paper E).

Direction 2.2: *Identify and study successful configurations of conversational user interfaces.* In this dissertation, a behavioral perspective was applied to study the effects of selected design features and configurations on selected user experiences and service outcomes. However, CIs afford a myriad of different design features, increasing the complexity of configurational IS design (Feine et al., 2019; Knote et al., 2021). Moreover, the effectiveness of specific CI design can only be determined by considering contextual variables, such as user demographics and service tasks. To furnish such a holistic picture, future research should adopt a configurational perspective on CI design. Specifically, future research could apply methodologies using set theories, such as qualitative comparative analysis, to identify and explain successful assemblages of CI design features for different application contexts (Paper A).

Direction 2.3: *Develop methods to systematically design conversational service interactions.* The body of knowledge on conversational interfaces, including this dissertation, offers a multitude of insights for the theory-driven design of conversational service interactions. However, owing to the high degree of complexity of conversational service interactions, it is necessary to arrive at methods that support the design of conversational service interactions across service tasks and contexts. For instance, several low-code platforms were created that support service firms in creating conversational service interaction following a guided process. However, these platforms often focus on supporting practitioners in deploying low-level implementations and offer few guidelines for successful conversational design (Winkler et al., 2020). Future research should therefore develop methods and approaches for the systematic design of conversational service interactions incorporating both theory- and user-driven insights into conversational design.

Direction 2.4: *Investigate designs of human-in-the-loop approaches for automated conversational service encounters.* In this dissertation, only fully automated service interactions were considered. However, owing to user requirements and the technologies' degree of maturity, service requests are often processed in a tandem consisting of a CI and a service employee. For example, the CI assumes the role of a porter, putting the user in touch with the relevant specialist department. These novel forms of AI–human teaming require substantial research efforts (Seeber et al., 2020). Specifically, future research could investigate how flow experiences might be maintained in conversational interactions that involve successive or overlapping interactions with both CIs and human service employees. To that end, researchers should investigate differential affordances

of humans and CIs for creating flow-like conversational user experiences and derive design principles for the handover between both actors.

5.5.3 Avenue 3: Potential of Voice-Based Automated Service Channels

This dissertation hints at the potential of novel service channels. Specifically, the unique affordances of voice-based CIs (see Paper E) provide new modes of service delivery. Moreover, they allow to collect vocal data that can potentially be used to derive rich user insights and to act on these insights to enhance service outcomes. However, research on when and how voice-based technologies can best be employed is scarce. In particular, three research directions emerge:

Direction 3.1: *Study how voice analytics can be successfully deployed across the service value chain.* Another promising area for future research lies in the application of voice analytics to derive unique and rich user insights. As described in the previous section, voice-based interactions permit the collection of vocal data that are richer than textual data. As such, the human voice incorporates various subtle cues that can be used to draw inferences about users (Hildebrand et al., 2020; Sutton et al., 2019). Service firms can use this information in a variety of ways. Specifically, voice analytics could enable firms to monitor service satisfaction in automated interactions or to evaluate the veracity of communicated messages as a tool for fraud detection. Moreover, vocal data may be used to create new service offerings, such as capturing users' vocal profiles of users for the purpose of customization and identification (Hildebrand et al., 2020). Hence, future research could identify and systematically assess the application potential of voice analytics across different contexts.

Direction 3.2: *Explore for which service tasks a voice interface is a suitable addition to a multi-channel offering.* In the future, companies will increasingly have to choose between different service channels for different service tasks to build a compelling multi-channel offering (Lehrer & Trenz, 2022; J. Singh et al., 2020). In this dissertation, only a limited number of tasks (i.e., filing an insurance claim or obtaining a product recommendation) were considered to study the effect of voice-based CIs. However, the use of a specific interface modality may not be beneficial across all tasks and contexts, as this dissertation shows. Moreover, it is important to examine the potential of voice-based CIs against other forms of service provision, such as interpersonal forms of customer service. Future research should thus explore the conditions under which voice-based CIs can enhance user experiences and when they become detrimental (Paper E).

Direction 3.3: *Explore when and how the sharing of vocal (as opposed to textual) data affects user experiences and service outcomes.* According to the conceptualization detailed in Paper E, vocal data differ from textual data with respect to the degree of media richness. As such, the sharing of vocal data may fundamentally affect the user experience (e.g., by increasing privacy concerns; Hildebrand et al., 2020). However, research exploring how users feel about sharing vocal data and whether differences in vocal vs. textual data affect sharing behavior is scarce. Future studies should also address the downstream consequences of service firms using vocal data. Firms providing services based on vocal data (e.g., voice authentication) may conceivably be perceived as more innovative (Hildebrand et al., 2020). However, the use of vocal data may also create new vulnerabilities relating to users' privacy concerns. Specifically, security breaches involving vocal data may be perceived as more serious by users and may thus have a more immediate and lasting impact on firm evaluation (Hildebrand et al., 2020).

5.5.4 Avenue 4: Ethical Implications of Gendered Interfaces

The results of this dissertation indicate that anthropomorphized interfaces can evoke gender stereotyping. However, few studies have adopted an ethical perspective on studying the impact of such processes on user experiences and service outcomes (Diederich et al., 2022; Sutton, 2020). Considering the prevalence of gendered interfaces and the lack of research on the ethical implications of such designs, the following two research directions are proposed:

Direction 4.1: *Analyze when and how anthropomorphized interfaces affect users' mental models and the adverse effects on users' attitudes and behaviors.* The results of this dissertation indicate that gender-specific cues in CIs can evoke gender-inspired social categorizations. Despite the many positive aspects of anthropomorphism, these cues can unwillingly and willingly create "false" mental models regarding these interfaces, ultimately leading to negative user outcomes (Wambsganss et al., 2021). Specifically, imbuing CIs with anthropomorphic cues could provoke users to make riskier decisions when interacting with service firms. For instance, Hildebrand & Bergner (2020) show that conversational (as opposed to static) interfaces can motivate users to make portfolio decisions that are significantly beyond their risk level. Hence, future research should investigate how the salient cues provided by CIs influence users' perceptions of these interfaces and what the downstream consequences are for users in critical and sensitive service contexts. The results of this dissertation suggest that gender-specific cues may be a good starting point for such investigations.

Direction 4.2: *Study unintended downstream consequences of gendered interfaces toward interpersonal interactions and identify ways to mitigate these.* This dissertation's results indicate that users apply gender-inspired social heuristics to evaluate CIs. However, some questions remain regarding the effect of gendered interfaces. It is unclear whether gendered interfaces reinforce gender stereotypes, as is often feared (UNESCO, 2019). Hence, future research could investigate whether interactions with gendered interfaces result in spillover effects to human interactions. Specifically, experimental designs that expose users to faulty interfaces could be used as stereotypical associations often occur in situations of stress and uncertainty (Waytz et al., 2014). Moreover, the question arises about how the formation of gender stereotypes may be prevented. Based on this study's results, the potential of gender-ambiguous voices could be further explored (see also Paper C).

Overall, these research opportunities represent an ambitious agenda that aims to advance research on CIs in service contexts. Overall, this agenda will inspire researchers to create insights that support a more human-centric future for service delivery in an increasingly AI-driven economy.

Part B: Papers of the Dissertation

Part B consists of the six research papers and corresponding bibliographical information. The papers have been re-formatted to ensure consistency between them. To that end, figures and tables have been numbered across the whole document, font styles and sizes have been aligned, and the references have been merged into one document. However, in line with the university regulations, the text of the papers was not altered. Thus, there are some inconsistencies regarding the terms used.

Paper A – A Review of the Empirical Literature on Conversational Agents and Future Research Directions

Table 7. Bibliographical metadata Paper A

Title	A Review of the Empirical Literature on Conversational Agents and Future Research Directions
Authors	Zierau, Naim; Elshan, Edona; Visini, Camillo; Janson, Andreas
Publication outlet	Proceedings of the 31st International Conference on Information Systems (ICIS 2020), Hyderabad, India.
Publication type	Conference Paper
Publication year	2020
Rating (Jourqual 3)	A
Publication status	Published

Abstract

The knowledge base related to user interaction with conversational agents (CAs) has grown dramatically but remains segregated. In this paper, we conduct a systematic literature review to investigate user interaction with CAs. We examined 107 papers published in outlets related to IS and HCI research. Then, we coded for design elements and user interaction outcomes, and isolated 7 significant determinants of these outcomes, as well as 42 themes with inconsistent evidence, providing grounds for future research. Building upon the insights from the analysis, we propose a research agenda to guide future research surrounding user interaction with CAs. Ultimately, we aim to contribute to the body of knowledge of IS and HCI in general and user interaction with CA in particular by indicating how developed a research field is regarding the number and content of the respective contributions. Furthermore, practitioners benefit from a structured overview related to CA design effects.

Keywords

Conversational Agent, Human-Computer Interaction, Literature Review,
Research Agenda

A.1 Introduction

Recently, there has been a steep increase in the interest of researchers and practitioners into Conversational Agents (CA) (Pfeuffer, Benlian, et al., 2019). These agents assist by engaging with users via natural language (Janssen et al., 2020). Thereby, they herald an enormous potential for digital disruption both for company-based processes (i.e., through cognitive automation) and user-based processes (i.e., through humanization of human-computer interaction; Maedche et al., 2019). In regards to the latter, these agents represent a novel form of Information Systems (IS) that can be distinguished from other entities of IS by their high degree of interaction and intelligence (Maedche et al., 2019). These capabilities may fundamentally affect user perceptions and raise a number of theory- and design-related questions, most prominently revolving around an emergent conversation-based interaction paradigm (e.g., L. Clark et al., 2019). Hence, the transition to CAs amplifies several challenges and calls for future research (Pfeuffer, Benlian, et al., 2019).

Thus, in recent years a diverse body of empirical work has emerged on CAs in different disciplines most prominently in the IS and Human-Computer-Interaction (HCI) domains (Janssen et al., 2020). Thereby, researchers have studied CAs from a variety of user interaction outcomes (i.e., attitudes, perceptions, intentions, behavior). Moreover, they have explored the effects of a multitude of design elements offered by those interfaces on these interaction outcomes (e.g., Feine et al., 2019; Janssen et al., 2020). Thus, building upon extant theories and reflecting the complex nature of CAs, researchers to date have analyzed a growing number of relationships between independent and dependent variables (DV). However, the heterogeneity among variables investigated in these studies leads to difficulty in summarizing, analyzing, and evaluating findings succinctly from the overall body of empirical research on CAs. This leads to a fragmented literature base and sometimes contradictory research results as this paper shows.

Although initial literature reviews and meta-studies (i.e., taxonomic classifications on CAs; Janssen et al., 2020) emerged during the past years, the research is still scattered across different streams of research. For instance, existing literature reviews focus on specific sub-classes of CAs, such as pedagogical agents (Winkler & Söllner, 2018), or generally discuss literature on AI-based applications (Rzepka & Berger, 2018), while classification-based papers focus on the structural characteristics of CAs (Janssen et al., 2020). Thus, the scientific and practical knowledge that has grown dramatically in recent years as this review shows remains segregated. However, relatively young research

fields such as the research on CAs need to arrive at an integrated conceptualization and synthesis of representative literature on which future research efforts can build on (Torraco, 2005). So far, to the best of our knowledge, such an integrated conceptualization does not exist, which results in research on CAs being terminologically fuzzy. Thus, we aim to encapsulate the rapidly developing empirical body of knowledge on CAs in a way that is concise, meaningful and provides value to researchers. Hence, this paper focuses on answering the following research question (RQ):

RQ: *What is the state of the art regarding user interactions with Conversational Agents?*

To answer this RQ, within the scope of this review, we examined 107 empirical publications on CAs in 20 seminal outlets related to IS and HCI research. We analyzed a multitude of findings across a selection of studies and aggregated the results from both quantitative and qualitative research. Based on the review, we extracted the most frequently studied constructs and developed four descriptive models pertaining to major user interaction outcomes. These models serve as a representation of the current state of the art on CA research, and the resulting gaps identify potential research avenues.

A.2 Conceptual Background

In recent years, scientific and industry interest in CAs has grown considerably (e.g., Feine et al., 2019; Pfeuffer, Benlian, et al., 2019). The basis for the new technology was laid in 1966 when Joseph Weizenbaum developed a computer program that communicated with humans via a text-based interface and which passed the touring test (Weizenbaum, 1966). These text-based interfaces were later followed by the development of voice-based dialogue systems and embodied conversational agents in the 1980s (McTear et al., 2016). The increased interest in this system type results from a few overlapping trends. On the one hand, due to recent advances in AI, especially in regards to natural language processing, new generations of CAs emerged that can be used to automate an increasing number of tasks in areas such as health (Laumer et al., 2019), education (Winkler et al., 2019), and customer service (Qiu & Benbasat, 2009).

On the other hand, the conversational character of CAs enables new as well as potentially more convenient and personal ways to access content and services, thus ultimately enhancing user interactions with IS. Along with these developments, the scientific interest in how these interfaces affect user perceptions has also increased. Under the terms of Intelligent Personal Assistant (Hauswald et al., 2016), Smart Personal Assistant (Knote et al., 2018), Chatbot (Følstad & Brandtzaeg, 2017), and Conversational Agent

(Feine et al., 2019), numerous studies have been undertaken in recent years. Here, we will highlight some key features of CAs as the overarching object of interest of this paper.

According to Maedche et al. (2019), CAs are distinguishable from other entities of IS based on their capabilities for interaction and intelligence. Regarding the first dimension, the ability to engage with users via natural language is formative to our understanding of CAs (Feine et al., 2019). Typically, CAs have relied on rigid behavioral patterns. Those agents could only respond to simple requests by matching user inputs against a set of stored patterns (McTear et al., 2016). However, novel forms of CAs can now process compound natural language and thereby respond to increasingly complex user requests (Knote et al., 2018). One example is Amazon's Alexa, which supports users to carry out everyday tasks via an advanced voice interface, ultimately acting as their personal assistant (Benlian et al., 2020). These agents increasingly mimic human-to-human interaction (Feine et al., 2019; Purington et al., 2017) and enable a more convenient and natural way to interact with technology (Knote et al., 2019). In addition, modern CAs are now usually also characterized by an intelligent component (Maedche et al., 2019). This property makes CAs more adaptive to different users and given context situations. Thus, CAs are capable to "learn" using inputs such as environmental data and user preferences (Maedche et al., 2019). Building on a continuously growing data set, CAs can adapt and personalize their behavior over time, thus manifesting autonomous characteristics (Pfeuffer, Benlian, et al., 2019). We consider a broad variety of CAs incorporating also less advanced agents (i.e., rule-based or scripted CAs) to give a comprehensive overview of respective user interactions.

In sum, these capabilities may fundamentally affect user interactions with these systems and raise several questions related to the theoretical grounding and design elements of CAs. In this regard, it was shown that the human-like traits of CAs might trigger users to exhibit emotional, cognitive, or behavioral reactions that are reminiscent of human interactions (Krämer et al., 2005). Hence, researchers are increasingly relying on the Computer Are Social Actors (CASA) paradigm as their theoretical foundation to explain specific user interaction outcomes. Accordingly, humans perceive certain CA design elements (e.g., an avatar) that cause them to categorize a technical system as a relevant social entity (Nass et al., 1994). In this regard, design elements can be defined as the distinctive technical, contextual, and knowledge features that frame the CA (Janssen et al., 2020). Recently, the inventory of possible design elements both concerning verbal and non-verbal communication has been significantly increasing (Feine et al., 2019),

which has enabled CA developers to address prevalent user concerns (i.e., lack of trust) and design increasingly convincing user interaction experiences (Pfeuffer, Benlian, et al., 2019). At the same time, a multitude of research emerged in different disciplines, most prominently in the IS and HCI domain, which empirically investigated the effect of specific CA design elements on various interaction outcomes. Thereby, most studies focused on single or few design elements or configurations and their effect on selected user perceptions, leading to a fragmented literature base, and sometimes contradictory research results.

This shortcoming could be addressed by an integrated analysis aggregating empirical insights on the diversity of CA design elements, which would increase our understanding of user behaviors and support us in identifying future research needs. Existing reviews on CAs either assume an overall perspective on AI-based technologies (e.g., Rzepka & Berger, 2018), which seems arguably too broad to draw meaningful conclusions from user interaction based on the specific characteristics of CAs. Alternatively, they focus on specific domains such as education (e.g., Winkler & Söllner, 2018), which is too narrow to draw overall conclusions on user interactions based on the breadth of application contexts and possible design elements (Pfeuffer, Benlian, et al., 2019). Moreover, to the best of our knowledge, there exists no review that takes a distinct perspective on the empirical effects of CAs despite the accelerating and at the same time segregated manner in which practical and scientific knowledge in this area has been growing (Janssen et al., 2020). Hence, we address the lack of an integrative perspective by conducting a systematic analysis of the empirical literature on CAs, to identify validated findings and reveal relevant research needs.

A.3 Research Approach

Hereafter, we describe our research approach to review current empirical CA literature, which was informed by the methodological approach employed by Jeyaraj et al. (2006). To that end, we followed the steps of identifying, coding, validating, and analyzing quantitative and qualitative empirical findings.

A.3.1 Paper Selection Process

To identify relevant literature as the basis for the state-of-the-art analysis, we conducted a Systematic Literature Review (SLR) following Webster & Watson (2002) and vom Brocke et al. (2015). The overall scope of the conducted SLR can be defined along the dimensions of process, source, coverage, and techniques of the SLR (vom Brocke et al.,

2015): Based on a sequential search process we searched relevant journals and conference proceedings from the field of IS and HCI literature as a source. Thereby, our literature search intends to reach a representative coverage of the design elements reported in the literature. Thus, to establish the basis for our analysis, we used a comprehensive set of techniques (i.e., keyword search, backward and forward search). To reach a high level of reproducibility and transparency of our research, we will describe the three single methodological steps that we undertook.

In the first step, we selected the search strings. Since we aimed to identify a wide range of literature on CAs, the search string is chosen to be rather broad. Based on recent publications, we identified different keywords researchers used to describe CAs. This resulted in the following search string: (“conversational agent” OR “chat bot” OR “chatbot” OR “dialogue system” OR “smart personal assistant” OR “smart assistant” OR “intelligent assistant”). In the SLR, we used all variations of the keywords – singular, plural, hyphenated, or not hyphenated. In the second step, we selected the outlets. As our goal is to identify representative literature samples of different empirical research perspectives on user interaction with CAs, our search covers multiple journals and conference proceedings. We choose this approach since journal acceptance processes take substantially longer than conference proceedings to be processed, which would lead to neglecting some of the most relevant literature since CA research represents a young and nascent topic. For the selection of outlets, we identified two broad areas for deriving design elements of CAs – IS and HCI – as they cover a substantial share of literature on CAs.

Suitable journals and conference proceedings at the intersection of HCI and IS that provide an overview of high-quality and relevant research in the respective research fields were selected using the Basket of Eight and relevant IS journals and conferences based on the recommendations of the Special Interest Group on Human-Computer Interaction. Moreover, to safeguard the relevance of our results, we discussed our selection of journals and conference proceedings with two senior researchers from the field of interest, who were not involved in the writing process of the paper. Based on these inputs and their feedback, we selected 20 journals and proceedings for our keyword search, as seen in Table 8. Finally, in the third step, we identified suitable publications for our review. Searching in the title, abstract, and keywords of the papers, the outlet-based search reveals 383 hits. This number still contains literature not relevant to this paper. In an initial screening process, the identified papers are analyzed based on their abstracts. We only included papers that referred to any type of CAs and which provide empirical insights

on user interaction with CAs. Papers dealing with this topic trivially or marginally, such as those generally dealing with technology acceptance of CAs, were removed from the sample. This resulted in a selection of 76 publications. Finally, we performed a forward and backward search to also capture papers not covered through the database search. Through screening the references and applying forward searches using Google Scholar, 31 articles were added to the list, including 24 papers from outlets that were not considered in the initial search, resulting in the final number of 107 relevant papers.

Table 8. Overview of searched journals and conference proceedings

Field	Outlets (<i>Hits/Relevant Publications/Additional Back- & Forward Searches in Brackets</i>)
IS	ACM Transactions on Information Systems (16/0/0), Decision Sciences (6/0/0), Decision Support Systems (39/5/1), European Journal of Information Systems (6/0/0), Information Systems Journal (2/0/0), Information Systems Research (6/0/0), Journal of Information Technology (1/0/0), Journal of Management Information Systems (21/1/0), Journal of Strategic Information Systems (0/0/0), Journal of the Association for Information Systems (1/0/0), Management Information Systems Quarterly (0/0/0), Proceedings of the International Conference on Information Systems (ICIS) (14/2/1). <i>Proceedings of European Conference on Information Systems</i> (3/0/2)
HCI	ACM Transactions on Computer-Human Interaction (17/6/0), Human-Computer Interaction (12/3/0), International Journal on Human-Computer Studies (43/6/0), Journal of Computer-Mediated Communication (6/0/0), Journal of the ACM (1/0/0), User-Modelling and User-Adapted Interaction (17/2/0), Proceedings of the Conference on Human Factors in Computing Systems (172/51/3), <i>Additional Back/Forward</i> (24)

A.3.2 Paper Analysis

The 107 relevant papers are analyzed from a concept-centric perspective using an abductive approach. Thereby, to aggregate the insights from identified studies, we developed a list of master codes and master code descriptions. This process was iterative and required multiple coding rounds of the identified papers by different researchers. Thereby, the iterative process started by two of the researchers to independently code a

subset of 20 randomly chosen articles. For each of the 20 studies, we listed each dependent and independent variable (IV) as named by the author(s), which together form our initial list of “author variables” and “author variable descriptions”.

Moreover, we captured contextual variables such as the application domain and task of the CA. Subsequently, we discussed how to combine variables across studies, which resulted in a list of “master variables” and “master variable descriptions”. In some cases, both lists are identical, while other variables required further consideration. Next, we re-examined the initial subset set of 20 articles and mapped author variables to our master variables. During the next iteration, two researchers independently coded another subset of 20 articles. Thereby, we coded the dependent, independent, and structural variables and also mapped these variables to the growing list of master variables and descriptions. Afterward, these researchers discussed their independent findings. In case the respective findings differed, a third researcher, was involved in discussing the differences. For each iteration, we added new master variables and descriptions. As new variables emerged, we reviewed previously coded articles to determine if any needed to be refined based on the addition of new master variables. This process was concluded once all articles were coded.

Concurrent with the creation of a list of master variables and descriptions, we also coded for empirical relationships between an independent and a DV in each study. Thereby, following Lacity et al. (2010), we assigned four possible codes to the relationship between independent and DVs: “+1”, “-1”, “0” and “M”. In this process, we coded “+1” for a positive relationship, “-1” for a negative relationship, and “0” for relationships that were studied but did not show any significant value in the empirical results. In quantitative studies, we used $P < 0.10$ as the requirement for a significant positive or negative relationship. In case the study was qualitative, we relied on the authors' argumentation signified by a robust theoretical anchoring, which we coded as “M”. All told, we coded 302 relationships between independent and DVs. Of these, 138 were positive and significant, 33 were negative and significant, 36 relationships mattered, and 95 relationships were not significant. The overrepresentation of positive effects on our DVs can either be explained by a strong focus on positive effects, or by the resistance of researchers to report negative effects regarding user interaction with CAs.

A.4 Results

We organized the findings into four sections. The first section examines descriptive statistics based on the meta-data of the underlying literature, the second section examines

the DVs used in CA research, the third section examines the IVs used in CA research, and the fourth examines the relationships between independent and DVs.

A.4.1 Section 1: Descriptive Findings

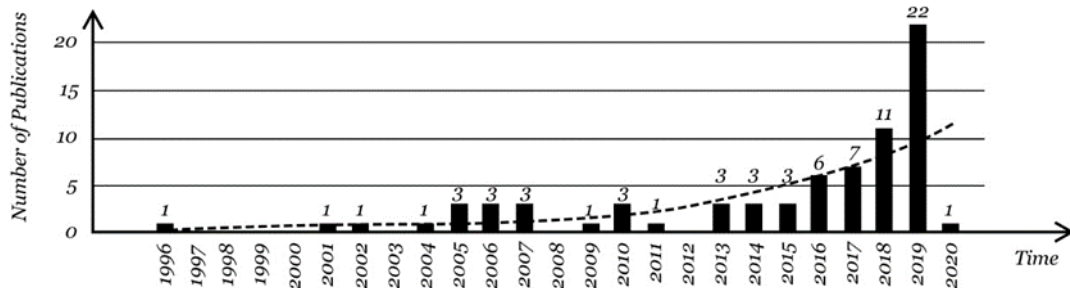


Figure 6. Number of publications over time (n=74, included in analysis of findings)

Figure 6. shows that the number of identified publications has been steeply growing during the last years. The youngest paper is from 2020 and the oldest paper from 1996. The majority of papers have been published within the last four years, which supports our initial assumption that CAs represent an emerging research field. This is also reflected by the fact that most papers are from conference proceedings, which gives testament to the relative youth of the field. Moreover, it is worth noting that a multitude of investigated papers is from the HCI discipline, while publications in IS outlets are only recently directing attention towards CAs. The first contributions are rather explorative, incorporating a multitude of investigated variables, while recent papers are more specific concerning the theoretical lenses applied and the effects investigated. Further, the examined contributions included empirical data from various application contexts and data sources.

A.4.2 Section 2: Findings on Dependet Variables

The publications at hand adopted a wide dispersion of DVs. We identified 390 DVs. We categorized these 390 DV into five broad categories: DVs that investigated perceptual and attitudinal outcomes. Moreover, one category summarizes behavioral intentions and outcomes, which represents a broad umbrella term based on a scarcity of findings in this regard (25 findings.). Thus, the five major variables that evolved from our review are rapport, social presence, trust, and utility as well as behavioral intention and outcomes.

Rapport (31.3%). Researchers have generally examined a plethora of outcomes related to the quality of the social bond between the user and the CA, which is also referred to as rapport (Pecune et al., 2018). A third of all studied outcome variables can be assigned to this category. Prominently studied variables in this category are the likeability of the

CA (e.g., Chin & Yi, 2019; Miehle et al., 2018) the degree of involvement or engagement experienced by users (e.g., Van Es et al., 2002; Van Vugt et al., 2010), and the perceived closeness (T. W. Bickmore & Picard, 2005; S. Y. Lee & Choi, 2017).

Social Presence (16.4%). Another important outcome category represents social presence, which does not seem surprising since many researchers work on recreating human-CA interactions that are experienced as human-like. In this regard, social presence can be defined as “*the extent to which other beings in the world appear to exist and react to the user*” (Heeter, 1992). Within this category, researchers focused on perceived humanness (e.g., Candello et al., 2017) and social presence (e.g., E. Cho, 2019) as the two main outcome variables.

Trust (15.2%). Additionally, a major perceptual outcome category is reflected by user trust. As many researchers cite a lack of trust as one of the central adoption barriers for AI-based technologies, this sentiment has also been important to trust researchers in regards to CAs making trust as one the main variables being in the focus of CA research (e.g., Kang & Wei, 2018). Trust is usually defined as an expectation that another entity “*will perform a particular action important to the trustor [i.e., user], irrespective of the ability to monitor or control that other party [i.e., CA]*” (R. C. Mayer et al., 1995, p. 712). However, authors also investigated trust-related concepts such as credibility (e.g., (Cowell & Stanney, 2005) or privacy perceptions (e.g., Benlian et al., 2020), which we incorporated in this section.

Utility (31.1%). A multitude of authors investigated productivity-related perceptions, which we summarized under the category of utility. Thereby, prior researchers have looked at usefulness (e.g., Qiu & Benbasat, 2010), ease of use (e.g., Van Es et al., 2002), the quality of interaction (e.g., Ashktorab et al., 2019) satisfaction (e.g., Chaves & Gerosa, 2018), and helpfulness (e.g., Berry et al., 2005).

Behavioral Intention and Outcomes (6.0%). Finally, few authors also investigated behavioral intentions and outcomes, whereby actual outcomes were in the minority. In regards to behavioral intention, prior authors have looked at the intention to use (e.g., Wunderlich & Paluch, 2017) and willingness to interact (e.g., Candello et al., 2017). Behavioral outcomes incorporated findings related to interaction behavior (e.g., length of interaction; (Strait et al., 2015) or behavior change (e.g., exercise behavior; T. W. Bickmore & Picard, 2005).

A.4.3 Section 3: Findings on Independent Variables

We identified 390 IVs used in CA research. To facilitate the discussion of this high quantity of IVs, we categorized them into five broader categories. Thereby, our allocation into aggregated dimensions is based on a taxonomic classification of social cues of CAs introduced by (Feine et al., 2019), which we extended based on our coding by selected categories, as some design elements did not fit these categories (i.e., interaction). Each category is briefly discussed below.

The category Auditory (3.3 %) refers to all design elements that can be perceived via the sense of hearing except the words itself (Burgoon et al., 2013). Within this dimension, those elements have been analyzed that relate to voice qualities representing permanent and adjustable characteristics of the voice. In total, these cues have been investigated 10 times. For example, Yu et al. (2019) have studied the impact of the voice's gender (female vs. male) on different perceptual outcomes. Although this category also hypothetically includes nonlinguistic vocals and sounds, there were no studies in our sample addressing these elements.

Within the category Interaction (14.5 %), we summarize all design elements that refer to the underlying structural representation of the interaction both in regards to its communication mode and its turn-taking mechanism. Overall, researchers have studied often the choice of interaction mode. Moreover, researchers studied the influence of preset answers, which reflects the degree of freedom employed in the conversation. The former category has been studied 34 times, whereby most researchers compared chat and voice interfaces (S. Kim et al., 2019). In comparison, the latter category was studied less but was found equally influential for user perceptions (Diederich et al., 2019).

Design elements in the category Invisible (10.6 %) refer to all CA characteristics that cannot be perceived by the sense of hearing or seeing (Knapp et al., 2013). These elements are often referred to as “the silent language” (Hall, 1990). This category can be divided into four subcategories: Chronemics refers to the role of timing in conversation and is reflected in studies that focus on the design of conversation flows (e.g., Winkler et al., 2019), system response times (e.g., Gnewuch et al., 2018) or the role of synchronicity (e.g., Park & Sundar, 2015), which in total have been studied 11 times. Intelligence refers to elements that express the cleverness of the agent, which was exemplarily studied by Xu et al. (2017). The other two categories, personality (i.e., personality traits) and haptics (i.e., tactile sensations), were comparatively less frequent in the research field. Among personality, e.g., Cafaro et al. (2013) examined the personality traits.

The category Visual (34.3 %) refers to all nonverbal design elements that are not invisible, and can visually be perceived except words itself (Leathers & Eaves, 2015). This category can be distinguished into five subcategories, which have been broadly studied (104 times). The most prominently researched variable is agent appearance (46 times). The embodiment of the agent has gained much attention in our sample (e.g., McBreen & Jack, 2001; Nunamaker et al., 2011). Another studied aspect of the agent's appearance was gender (e.g., Pfeuffer, Adam, et al., 2019). Furthermore, kinesics, which refers to body movements, such as demeanors (e.g., Krämer et al., 2013) and gaze patterns (e.g., Van Es et al., 2002) were addressed in total 25 times. Computer-mediated communication (CMC) refers to visual elements that augment or modify written texts and were examined 23 times. Here, the effects of using emojis (e.g., Park & Sundar, 2015), typos (e.g., Westerman et al., 2019) or videos and images (Huber et al., 2018) were researched. The other two categories, entrainment (i.e., the adjustment of visual elements to the user) and proxemics (i.e., the role of distance in communication), have not yet been studied in detail (10 times for both).

The category Verbal (37.3 %) refers to all CA elements that can be expressed by words either written or spoken (Antaki, 2008). Within this dimension, the conversation style, which refers to how something is being communicated, was the most researched IV (53 times). For instance, R. E. Mayer et al. (2006) studied the effects of relational strategies. The aspect of content captures all elements that relate to the literal meaning of a message and was researched a total of 22 times. For example, Akahori et al. (2019) have looked at the effects of self-disclosure. Similar attention has been given to adaptivity (22 times), which refers to the verbal adaptation of the CA to the users. Within this category, researchers have studied the use of contextual information (e.g., Vtyurina et al., 2017), user content (e.g., Schuetzler et al., 2014), or the absence of adaptivity (e.g., Engelhardt et al., 2017).

A.4.4 Section 4: Findings on the Relationship between Independent and Dependent Variables

In this section, we summarize major findings concerning the 49 relationships we coded between 16 IV and four DV (perceptual and attitudinal outcomes) as well as the regarding behavioral intention and outcomes (that we discuss more generally based on a scarcity of coded relationships). At this detailed level, the frequency with which findings were replicated across studies was minimal and did not provide a very coherent or comprehensive picture of CA research. Hence, to study these relationships in a way that is

concise, and helpful to researchers, we moved to a higher unit of analysis by reporting the 277 findings using our four categories of DV and the five categories of IV. Although precision is reduced in aggregating to the broader categories of DVs, we gain a better overall understanding of the determinants of perceptual and attitudinal outcomes of CAs. Thereby, we also aim to investigate the consistency of the empirical evidence.

For this reason, we indicated the consistency of the empirical evidence. Therefore, we used (!!) if the evidence was consistent in more than 80 percent of the reviewed relationships. Accordingly, if the relationship provided less robust results, we used a (!) evidence that reflected a consistency of 60 to 80 percent of the reviewed relationships. To indicate that a relationship did not meet the set thresholds, we indicated it with a (0). We then applied the same logic to relationships if the evidence originated from a qualitative study and was argumentative. Therefore, if the relationship was consistent in more than 80 percent of the reviewed relationships, we indicated it with a (MM). Accordingly, we applied the same logic and used an (M) if the relationship was less robust. For instance, if a relationship was studied six times, and one represented an insignificant relationship, and the remaining evidence showed a significant relationship, we categorized this example with a (!!).

Further, we included the experimental variables and marked them. Experimental variables refer to variables that have been studied less than five times. Those variables, as well as the inconclusive findings, guide in identifying avenues for further investigation and discussing perspectives for a research agenda.

Independent Variables on Rapport

Figure 7 provides an overview of the relationships between rapport and the five IVs. In this model, 92 findings are synthesized and depicted based on consistency within the subgroups of IVs, providing an answer to the question: “Which determinants of rapport were reported by past empirical research on CAs?”.

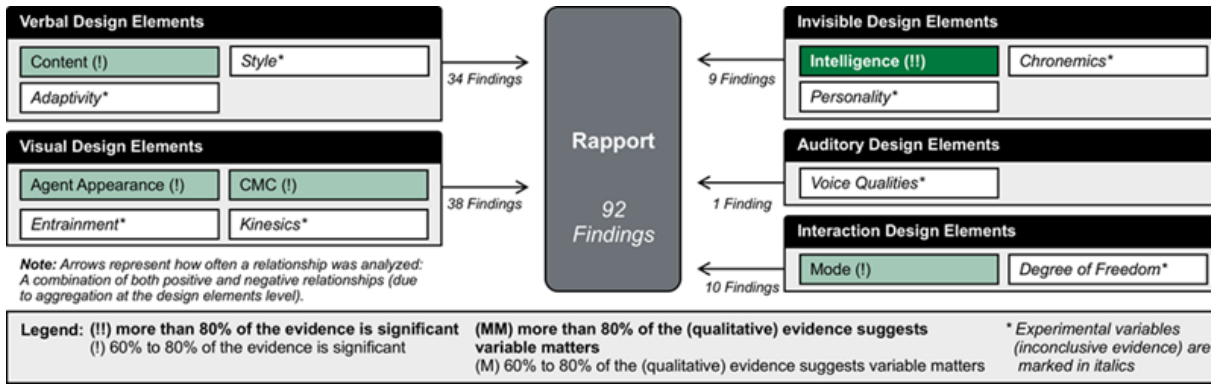


Figure 7. CA outcomes regarding rapport

In this model, we identified 1 variable showing consistency of 80% or more, as well as 4 variables showing consistent results in 60% of all instances. The other 8 variables were marked as experimental variables due to having less than 3 observed samples (4 variables) or showing inconsistent findings (4 variables).

Past research offers evidence that verbal design elements are an antecedent of rapport towards the CA. Researchers investigated the variables style (20 observations (OBS), e.g., L. Clark et al., 2019), adaptivity (10 OBS, e.g., S. Lee et al., 2019), and content (4 OBS, e.g., L. Clark et al., 2019). In this regard, content and entrainment were identified as having a significant impact. For instance, it was shown that eliciting similar interests (L. Clark et al., 2019) and the degree of matching or coordination in word counts of the CA and the user positively influences rapport-building (Pecune et al., 2018). Further, our findings indicate that visual design elements are determinants of rapport. Researchers have studied the variables agent appearance (19 OBS, e.g., S. Lee et al., 1996), kinesics (12 OBS, e.g., Krämer et al., 2013), entrainment (4 OBS, e.g., Qiu & Benbasat, 2010), and CMC (3 OBS, e.g., Westerman et al., 2019). Thereby, agent appearance and CMC were found to be significant. For instance, enriching the CA's message by the way of typos and capitalization uncovered a significant influence on the social attractiveness of the CA (Westerman et al., 2019). Moreover, including typos and capitalization as manifestations of CMC increased the social attractiveness of the CA (Westerman et al., 2019). In our model, also variables related to invisible design elements were found to be significant and consistent determinants of rapport. Researchers have inquired into the variables intelligence (6 OBS, e.g., Schuetzler et al., 2019), chronemics (2 OBS, e.g., Winkler et al., 2019), and personality (1 OBS, Cafaro et al., 2013).

In contrast, past research was found to have directed only limited attention to the influence of auditory design elements on rapport between the user and CA. Only one variable

(i.e., voice pitch) was studied (Yu et al., 2019), indicating no conclusive evidence. Additionally, our findings indicate some evidence regarding the influence of interaction design elements as determinants of rapport between the user and CA. Researchers studied two variables, mode (9 OBS, e.g., Miehle et al., 2018), and degree of freedom (1 OBS, Jeong et al., 2019). The influence of mode was found to be significant. For instance, employing voice-based interfaces increased users' self-disclosure towards the CA (Yu et al., 2019).

To summarize our findings on the DV rapport: The most significant and consistent evidence regarding determinants of this outcome dimension was found to be related to the group of variables coded as intelligence (invisible). Other consistent findings were found regarding the variables categorized as agent appearance and CMC (both visual), mode (interaction), as well as content (verbal).

Independent Variables on Social Presence

Our findings regarding the outcome dimension of social presence are outlined in Figure 8. In this model, 48 findings are synthesized and depicted based on their consistency within the subgroups of the five IVs, providing an answer to the question: “Which determinants of social presence were reported by past empirical research on CAs?”.

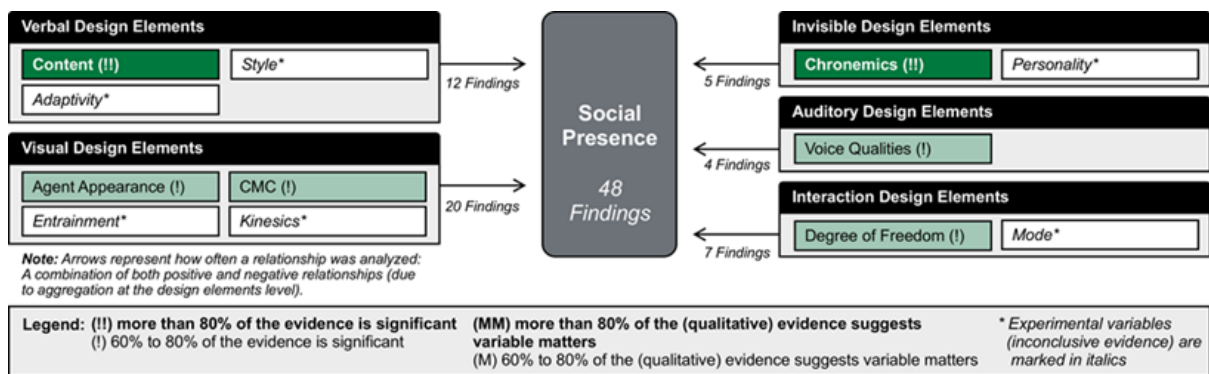


Figure 8. CA outcomes regarding social presence

In this model, we found 2 variables showing consistency of 80% or more, as well as 4 variables showing consistency of 60% or more. The 7 remaining variables were categorized as experimental variables due to having less than 3 observed samples (4 variables) or showing inconsistent findings (3 variables).

Past research offers some evidence that verbal design elements of conversational agents are determinants of social presence. Researchers have investigated the variables style (6 OBS, e.g., T. Bickmore & Schulman, 2007), content (3 OBS, Kobori et al., 2016), and

adaptivity (3 OBS, Schuetzler et al., 2014). Only content was found to be significant for social presence, i.e. small-talk utterances increased the perception of the liveliness of the agent (Kobori et al., 2016). Moreover, in our sample, we found considerable evidence of visual design elements being determinants of social presence. Researchers investigated the variables CMC (8 OBS, e.g., Candello et al., 2017), agent appearance (7 OBS, e.g., Lee et al. 2019), kinesics (3 OBS, Van Es et al., 2002), and entrainment (2 OBS, Qiu & Benbasat, 2010). For instance, a CA with a humanoid embodiment was found to be perceived as significantly higher in social presence as compared to a CA with no embodiment features. Additionally, our study uncovered considerable evidence suggesting that invisible design elements are determinants of social presence. Researchers have investigated the variables chronemics (3 OBS, e.g., Gnewuch et al., 2018) and personality (2 OBS, e.g., Q. V. Liao et al., 2018). For instance, dynamic delays in system response time, compared to near-instant responses, were observed to invoke higher perceptions of social presence and naturalness of the interaction (Gnewuch et al., 2018). Further, previous research on auditory design elements identified consistent and significant evidence on social presence (4 OBS, Qiu & Benbasat, 2009). Voice qualities was found to be a significant determinant of social presence. For instance, low pitch contour and high flanging increments were found to significantly affect perceptions of humanness (Muralidharan et al., 2014). In addition, past research studying interaction design elements on social presence identified consistent and significant evidence. Researchers have studied the variables mode (4 OBS, E. Cho, 2019) and degree of freedom (3 OBS, Diederich et al., 2019). Thereby, only variables related to the degree of freedom produced significant evidence. For example, pre-defined answer options were found to negatively affect perceptions of humanness (Diederich et al., 2019).

Summarizing our findings on the DV social presence, the most significant and consistent evidence regarding its determinants was found to be related to the two groups of variables coded as content (verbal), and chronemics (invisible). Other consistent and significant evidence was found regarding the variables agent appearance and CMC (both visual), voice qualities (auditory), and degree of freedom (interaction).

Independent Variables on Trust

Figure 9 provides an overview of the relationships between trust and the five IVs. In this model, 45 findings are synthesized and depicted based on their consistency within the subgroups of the IVs, providing an answer to the question: “Which determinants of trust were reported by past empirical research on CAs?”.

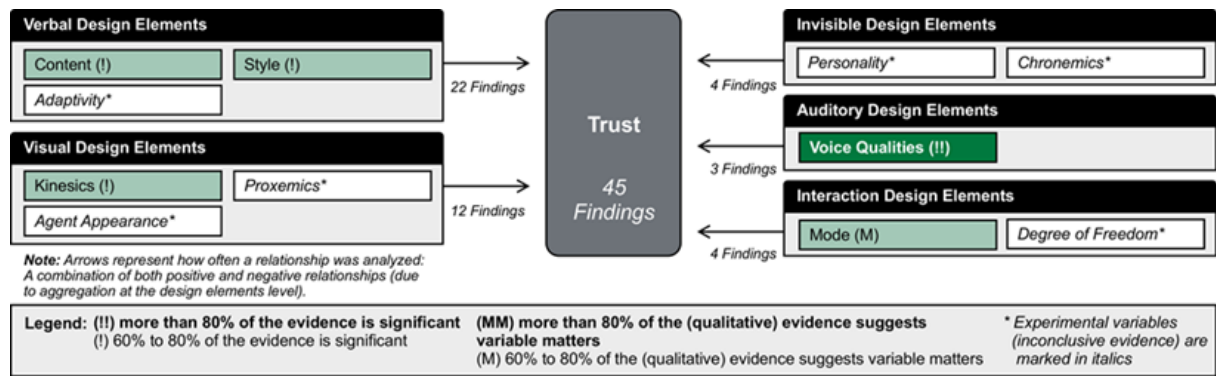


Figure 9. CA outcomes regarding trust

In this model, we found one variable showing consistency of 80% or more, as well as three variables showing consistency of 60% or more. Another six variables were categorized as experimental variables, due to having less than 3 observed samples (4 variables) or showing inconsistent findings (2 variables). The one remaining variable was categorized as argumentative.

Past research offers some evidence that verbal design elements are determinants of trust. Researchers have focused their investigation on the variables style (9 OBS, e.g., Kang & Wei, 2018), adaptivity (8 OBS, Engelhardt et al., 2017), and content (5 OBS, e.g., Benlian et al., 2020). The variable style produced nearly consistent results and was marked six times in a significant relationship. In contrast, the connection between adaptivity and trust showed mixed results. The variable content has, in turn, shown promising results. In this case, three findings were significant, and two more were investigated qualitatively. Contrary, past research offers little evidence that visual design elements of conversational agents are determinants of trust. Prior research has investigated the variables agent appearance (8 OBS, e.g., Nunamaker et al., 2011), kinesics (3 OBS, e.g., Elkins & Derrick, 2013), and proxemics (1 OBS, Benlian et al., 2020). Similarly, prior research has yet not been able to provide evidence that invisible design elements of conversational agents are determinants of trust. Nevertheless, past research has investigated the variables personality (3 OBS, e.g., Nordheim et al., 2019) and chronemics (1 OBS, Benlian et al., 2020). Concerning the variables personality and chronemics and their influence on trust, no conclusive findings have been elaborated. Additionally, there is little evidence that auditory design elements of conversational agents are determinants of trust. Nevertheless, past research has focused on voice qualities (3 OBS, e.g., Muralidharan et al., 2014) and found strong and consistent results. Moreover, prior investigations have not offered any evidence that interaction design elements are determinants of trust. However, the results obtained have a qualitative character.

Summarizing our findings on the DV trust, the most significant and consistent evidence regarding determinants of this outcome dimension was found to be related to the auditory design element of voice qualities. Additional evidence concerned the two groups of variables coded as style and content (verbal), mode (interaction), and kinesics (visual). Other variables have not yet been able to show significant evidence in relation to trust. However, there are some promising avenues for future research.

Independent Variables on Utility

Our findings regarding the outcome dimension of utility are outlined in Figure 10. In this model, 92 findings are synthesized and depicted based on their consistency within the subgroups of the IVs, providing an answer to the question: “Which determinants of utility were reported by past empirical research on CAs?”. We found three variables showing consistency of 80% or more, as well as three variables showing consistency of 60% or more. Another seven variables were categorized as experimental variables, due to having less than 3 observed samples (4 variables) or showing inconsistent findings (3 variables).

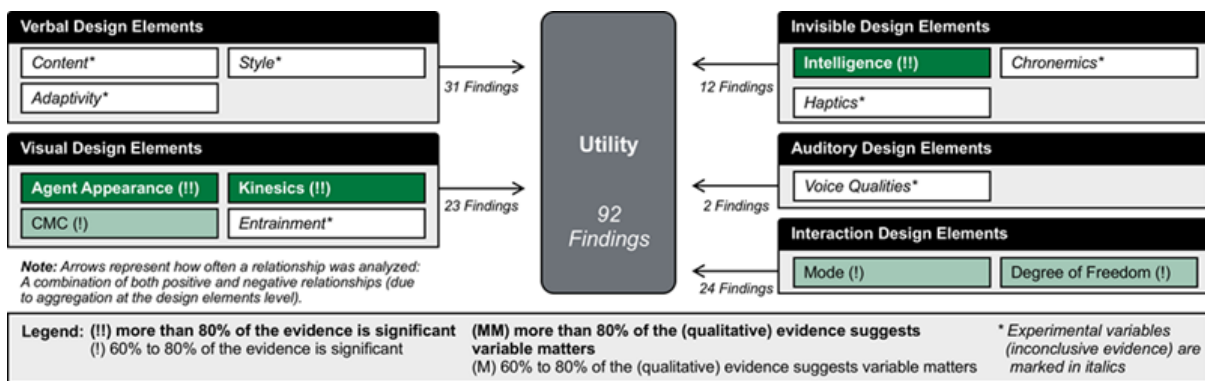


Figure 10. CA outcomes regarding utility

Past research offers no evidence that verbal design elements are determinants of utility. Researchers investigated the variables style (13 OBS, e.g., S. Kim et al., 2019), content (8 OBS, e.g., Kobori et al., 2016), and adaptivity (10 OBS, Engelhardt et al., 2017). The variables style, adaptivity, and content of the conversation have not yet been able to show evidence that the relationship to utility is significant. In contrast, we found considerable evidence of visual design elements being determinants of utility. Prior research has focused on the variables CMC (10 OBS, Westerman et al., 2019), kinesics (6 OBS, Cowell & Stanney, 2005), agent appearance (5 OBS, e.g., McBreen & Jack, 2001), and entrainment (2 OBS, Koulouri et al., 2016). For example, the embodiment of the CA with facial expressions was perceived as more useful, and the users seemed to be more satisfied than the faceless CA (S. Lee et al., 1996).

Users seemed to be more satisfied when the CA had a controlled but normal gaze pattern than when it had a randomized gaze pattern (van Es et al., 2002). Furthermore, past research offers evidence that invisible design elements are determinants of utility. Researchers have focused on the variables intelligence (6 OBS, e.g., Xu et al., 2017), chronemics (5 OBS, Chaves & Gerosa, 2018), and haptics (1 OBS, S. Kim et al., 2018). For instance, the perceived usefulness was higher, when the CA was empowered by deep learning, than when it did not (Xu et al., 2017). On the contrary, we found no evidence that auditory design elements are determinants of utility. Only the influence of voice qualities has been investigated, but no significant evidence was found (Tian et al., 2017). Voice qualities, i.e., the distinctive characteristics between acted and natural speech, did not affect how well the CA recognized the users' emotions.

Additionally, past research offers some evidence that interaction design elements are determinants of utility. Prior researchers have found significant evidence when looking at the variable mode (15 OBS, Miehle et al., 2018). Concerning the degree of freedom (9 OBS, Mu & Sarkar, 2019), no conclusive findings have been elaborated. For example, Akahori et al. (2019) were able to show that the main effects of the number of agents had a significant influence on understandability.

Summarizing the DV utility, the most significant and consistent evidence regarding determinants of this outcome dimension was related to the three groups of variables coded as intelligence (invisible), and agent appearance as well as kinesics (both visual). Other consistent and significant evidence was found regarding the variables CMC (visual), and degree of freedom as well as mode (both interaction).

Independent Variables on Behavioral Intention and Outcomes

Findings on the dimension of behavioral intention and outcomes are scarce (25 findings). Moreover, this category represents a broad category incorporating different intentions and behaviors, which is why we did not create an independent model that indicates the consistency within the subgroups of the IVs. In sum, past research mostly looked at visual (11 OBS; e.g., Strait et al., 2015) and verbal (10 OBS, e.g., Kobori et al., 2016) design elements. Only a minority of findings looked at interaction (2 OBS, e.g., Miehle et al., 2018) and invisible (2 OBS, e.g., Cafaro et al., 2013) design elements. No study in our sample investigated the effect of auditory design elements on behavioral intentions or outcomes. Summarizing our findings, we did not record any consistent findings within this subgroup. However, we find that while visual and verbal design elements are

starting to be investigated, others like auditory features have not yet attracted any research.

A.5 Discussion and Development of a Research Agenda

In the next paragraphs, we discuss the findings based on our research space. This is possible as both quantitative and qualitative assessments have been purposefully aggregated. The research agenda is positioned in the research space that we have spanned by this review linking important attitudinal outcomes of CAs to categories of design elements. Furthermore, we relate our results to the overall HCI framework. Table 9 provides a summary of the progress in various fields of research. The Harvey Balls indicate how developed a research field is regarding the number and content of the respective contributions. Fields where no research had been found, were classified as very low. Fields, in which only a few items were found or the contributions represent only a first attempt to research (i.e., only qualitatively), were classified as low. Fields with some contributions or mixed results were classified as moderate. Fields with several contributions and consistent findings were classified as high. Fields that have already been fully covered (very high) were not found. In connection with the descriptive and thematic findings, this assessment provides the opportunity to assume the need for further research. In each thematic discussion, research opportunities are localized and characterized. Thus, we follow the framework of Müller-Bloch & Kranz (2015). Also, practitioners get a structured overview of existing knowledge in the field of CA design.

Table 9. Summary of research process

Design Elements	Attitudinal and Perceptual Outcomes			
	Rapport	Social Presence	Trust	Utility
Interaction				
Visual				
Verbal				
Auditory				
Invisible				
Legend: Low  , Moderate  , High 				

The publications regarding rapport provide insights into design elements that contribute to forming a social bond with the agent. In total, we have found 92 relationships involving rapport within 40 unique studies, which testifies to the high research interest in this aspect of user interaction with the agent. Thereby, several relationships showed consistent and robust evidence of a positive effect on rapport. Notably, the CAs intelligence seems to afford a positive user evaluation of relationship to the agent (e.g., Xu et al., 2017). Moreover, we found consistent evidence that agent appearance and CMC (both visual) and content (verbal) may be positively related to rapport. However, there are still a few inconsistent findings, as shown in this review, which prompt further research. Potentially fruitful future research directions are particularly evident regarding the influence of the following design elements:

- **Mode (Interaction):** Understanding how rapport emerges between the user and CA in the context of different interaction modes has been the focus of several studies: In our review, we extracted 9 findings related to the influence of the interaction mode on rapport. For instance, D’Mello et al. (2010) investigated how the mode of interaction (text-based vs voice-based) influenced dynamics between students and a CA tutor, but the limited sample of the study moderates the explanatory power. Thus, we argue that exploring CA interaction mode and its influence on rapport represents a potentially insightful direction for future research.
- **Style (Verbal):** Across different studies, 20 findings were concerned with investigating the influence of the CA’s verbal style on rapport dynamics between the user and the CA. However, from an aggregate perspective, the results were mixed. Six findings reported inconclusive evidence regarding any relationship. Three studies argued a positive relationship but did not provide any quantitative evidence. Therefore, we identify rapport between the user and CA in the context of different CA verbal styles as a potentially worthwhile avenue for future research.

Regarding social presence, we identified 48 findings within 25 unique studies. Thereby, especially those design elements that represent social cues of CAs not surprisingly showed a strong and consistent effect on social presence. Typically, CA visual design is the primary research area when considering social presence. Though, the findings related to appearance are not as consistent as one would expect, especially since agent appearance is a well-researched area when considering early research on other agents (Nowak & Biocca, 2003). Thus, future research should incorporate a more nuanced and

configurational view on these design elements, since we suspect that the interrelationship could be key for understanding social presence with CAs. Hence, we propose the following prospects for future research:

- **Agent Appearance (Visual):** Future research should provide a more nuanced understanding of what appearance elements positively influence social presence and to what degree. Arguing from an “uncanny valley” perspective (Mori et al., 2012), specific elements of agent appearance could be related more consistently to leveraging social presence perceptions than others, especially anthropomorphic design elements (Pfeuffer, Benlian, et al., 2019).
- **Configurational View on Social Presence:** Connected with the first avenue for future research, we suggest that agent appearance from a visual perspective as well as other elements leveraging social presence should not be treated in isolation and carefully considered in a configurational view. For instance, with other aspects such as the personality of the agent, which has been, so far, neglected in research. For example, Amazon’s Alexa has a very minimalistic appearance but rather a high degree of social presence through other elements fostering social presence (Purinton et al., 2017).

The contributions regarding trust provide insights into design elements that form trust in the agent. In total, we have found 45 relationships involving trust within 23 unique studies. Thereby, several relationships showed strong evidence, which testifies to the relative maturity of this research stream. Especially, voice qualities seem to be strongly associated with trust-formation (e.g., Muralidharan et al., 2014). Moreover, we found consistent evidence that content and style (verbal), kinesics (visual), and mode (interaction) may be positively related to trust. However, these are results that should be corroborated by further studies and replicated in different contexts. Further research opportunities are especially evident regarding the points of:

- **Personality (Invisible):** The reasoning of Nordheim et al. (2019) indicates that attributing personality to CAs shall positively affect perceptions of trustworthiness. However, the papers in our sample show inconclusive evidence regarding the effects of agent personality. Hence, we propose that future research further explores the attribution of personality dimensions (i.e., Big Five) on CAs and their effects on trust.
- **Agent Appearance (Visual):** In general, agent appearance has been explored relatively often regarding user trust (e.g., Nunamaker et al., 2011). However, the findings only herald mixed results. It seems that agent appearance is vital in some

contexts and others not. Hence, there is a need to explore the effect of appearance design from a more nuanced perspective taking into account the respective context or task.

In the analyzed literature, manifold insights on how design elements contribute to creating utility for the user have been made. In this research stream, we identified 92 findings involving utility within 41 unique studies. Especially those characteristics afford utility that are concerned with the accessibility or functionality aspect of HCI. Thus, we found strong evidence for the effects of degree of freedom (interaction), intelligence (invisible), agent appearance, and kinesics (visual). Moreover, we found consistent evidence adaptivity (verbal), CMC (visual), and mode (interaction) may be positively related to utility; however, these are results that should be corroborated by further studies and replicated in different contexts. In general, the results regarding utility perceptions are already quite profound. However, we see merit, especially in the following research directions:

- **Adaptivity (Verbal):** Based on theoretical reasoning and qualitative data, several researchers highlighted that adaptivity might afford high potential for creating utility as users expect personalized content. However, other researchers argue that standardized content may contribute to ease of use (Chin & Yi, 2019). Hence, we propose to investigate the effect of CA adaptivity on utility in different contexts.
- **Mode (Interaction):** Researching the utility of interaction with CA is undoubtedly one of the most common research side effects of research in the field of CA. However, very few have been concerned with merely investigating the influence of the chosen interaction mode on utility.

Furthermore, we only found 25 outcomes related to intentions and behavioral outcomes. This represents another vital research gap as the discipline is in dire need to investigate the influence of all aspects of CAs and their features both on actual user behavior and on evoked behavior changes in real-life settings in various domains.

Finally, we have identified some overarching research opportunities based on the overall positioning of the reviewed literature. According to Li & Zhang (2005), HCI is concerned with the interaction between an IS and a user. This interaction is shaped by the characteristics of the system, the user, and the task and context. Interaction results can include perceptions, attitudes, intentions, and behaviors (Li & Zhang, 2005). In the scope of reviewing the literature, we found that task context (i.e., support, assistance,

coaching function) has been rarely implicated in the research design. Additionally, user characteristics have been hardly considered in the research model beyond being a control variable. Here, we see an important research gap as both task and user characteristics may dramatically influence the effect of CA design on user perceptions and behaviors.

A.6 Limitations

Despite us following established guidelines and attempting to rigorously analyze the identified empirical literature on user interaction with CAs, this SLR has several limitations that should be considered. First, the scope of this SLR is not exhaustive. Despite due diligence, the scope of the SLR might not be fully exhaustive, and our search strategy may have missed relevant publications. Nevertheless, we aimed at capturing a broad spectrum of research on user interactions with CAs by employing a journal as well as a proceedings-based search. Second, the indicated relationships between the design elements and user outcomes are based on our interpretation of prior empirical research. Furthermore, the number of findings ultimately coded and included in our dataset was limited. Thus, it is not our intention to suggest any kind of causality between the design elements and user outcomes. By employing the method introduced by Jeyaraj et al. (2006), it was our objective to elucidate the variables studied, and offer a conceptual structuring of the empirical findings on design elements and their influence on outcomes. Fourth, bias within the results was visible, which consisted of a strong overrepresentation of positive effects. Finally, the resulting research agenda imposes further limitations. As such, the resulting and presented research agenda cannot be regarded as complete. Additional research questions and streams might be formulated for each of the agenda's parts.

A.7 Conclusions

The holistic evaluation of the empirical academic literature regarding user interaction with CAs is crucial to uncover potential research fields and gaps for shaping future empirical CA research. For this purpose, we conducted a systematic literature review to study which design elements had a significant influence on design outcomes. Following Jeyaraj et al. (2006), we identified, coded, validated, and analyzed quantitative and qualitative empirical findings on user interaction with CAs. We, therefore, analyzed the 107 identified research papers and systematically identified existing knowledge as well as future research needs. Despite the already impressive growth of existing CA research,

the field continues to rapidly evolve. Hence, based on a systematic examination of relationships between major IVs and DVs, we have suggested areas of future research representing understudied or inconclusive yet promising variables and interactions.

.

Paper B – Trust in Smart Personal Assistants: A Systematic Literature Review and Development of a Research Agenda

Table 10. Bibliographical metadata Paper B

Title	Trust in Smart Personal Assistants: A Systematic Literature Review and Development of a Research Agenda
Authors	Zierau, Naim ; Engel, Christian; Söllner, Matthias; Leimeister, Jan Marco
Publication outlet	15th International Conference on Wirtschaftsinformatik (WI 2020).
Publication type	Conference Paper
Publication year	2020
Rating (Jourqual 3)	C
Publication status	Published

Abstract

Smart Personal Assistants (SPA) fundamentally influence the way individuals perform tasks, use services and interact with organizations. They thus bear an immense economic and societal potential. However, a lack of trust - rooted in perceptions of uncertainty and risk - when interacting with intelligent computer agents can inhibit their adoption. In this paper, we conduct a systematic literature review to investigate the state of knowledge on trust in SPAs. Based on a concept-centric analysis of 50 papers, we derive three distinct research perspectives that constitute this nascent field: user interface-driven, interaction-driven, and explanation-driven trust in SPAs. Building on the results of our analysis, we develop a research agenda to spark and guide future research surrounding trust in SPAs. Ultimately, this paper intends to contribute to the body of knowledge of trust in artificial intelligence-based systems, specifically SPAs. It does so by proposing a novel framework mapping out their relationship.

Keywords

Trust, Smart Personal Assistant, Conversational Agent, Literature Review, Research Agenda

B.1 Introduction

In recent years, technologies based on Artificial Intelligence (AI) have matured and are increasingly permeating the professional and private lives of many people (Maedche et al., 2019). A key area of application represents the proliferation of Smart Personal Assistants (SPA) – computer agents that provide assistance by engaging with users via natural language. SPAs, sometimes also referred to as intelligent agents or conversational agents, are now applied in a wide area of usage scenarios (Rzepka & Berger, 2018). These agents including Amazon's *Echo*, Google's *Google Assistant* and Apple's *Siri* fundamentally influence the way in which individuals complete tasks, search for information, consume product and service offerings, and interact with organizations (McLean & Osei-Frimpong, 2019). Thus, it is foreseen that SPAs will become a daily companion for a wide range of users – for instance, the number of worldwide users of these agents is predicted to rise to almost 1.8 billion in 2021, which reflects the social and economic impact SPAs exhibit on a global scale (Tractica, 2016).

However, the potential of SPAs can only be realized if users feel that they can trust the new medium (Cowan et al., 2017). Since SPAs rely on AI, they face corresponding problems in terms of user acceptance. In a wide range of domains, such as finance or medicine, the acceptance towards the recommendations of AI-based systems, is below 60 percent (Jensen et al., 2010). These findings are in line with a recent survey by Forbes which has shown that 41.5 percent of U.S. consumers do not trust any AI-infused digital assistants (Kroque, 2017). These numbers highlight that trust is paramount for helping users overcome adaption resistance due to perceptions of risk and uncertainty (McKnight et al., 2002). At the same time, however, the unique characteristics of AI-infused systems including their opaqueness, embedded biases, and autonomous nature may not only make it difficult to garner trust, but afford risks based on non-reflective reliance (J. D. Lee & See, 2004). Understanding the nature and antecedents of trust in intelligent agents such as SPAs as a combined unit of analysis is, therefore, a necessary requirement for both IS researchers and practitioners who aim to successfully design and deploy SPA-based systems. As we will show in this paper, there has been a steep increase in publications on the topic of trust in SPAs. For the progress of a young and emerging research field, it is important to present previous research coherently and transparently such that the important research streams are highlighted, and their interrelationships and theoretical basis are presented (Torraco, 2005). So far, to the best of our knowledge, an overview and conceptual structuring of the combined research field of trust and SPAs does not exist. This results in a lack of terminological clarity and theoretical integration

of important concepts. Our Systematic Literature Review (SLR) addresses this gap by contributing to creating a common language and structuring the conceptual basis of trust in SPAs. Thus, we intend to answer the following research question:

Which research streams conceptually constitute the research field of trust in Smart Personal Assistants?

Overall, this SLR intends to contribute to the body of knowledge of trust in AI-based systems in general and trust in SPAs in particular by providing an integrated theoretical framework of the latter. This framework suggests that trust in SPAs can be distinguished into three distinct research perspectives. Based on this conceptual grounding, we propose a preliminary research agenda to trigger and guide future research in this nascent field.

The remainder of this paper is structured as follows: First, we provide a brief summary of the theoretical background on both SPAs and trust. Subsequently, we conduct an SLR according to Webster & Watson (2002) and vom Brocke et al. (2009, 2015) to provide an overview and structuration of the field of trust in SPAs. Furthermore, we present an integrated theoretical framework incorporating all research streams within the joint research field of trust in SPAs that we could conceptionally derive from literature. Finally, based on the state of the art in the particular research streams a preliminary research agenda is proposed.

B.2 Theoretical Foundation

To provide a foundation for our integrative review and our discussion of our research agenda, we first begin by defining key concepts including SPAs and trust.

B.2.1 Smart Personal Assistants

Research on SPAs is not a new field per se. However, it recently gained wide prominence in the broader public. In the past, these systems were almost exclusively studied as expert systems giving “intelligent advice” within a limited set of highly specified use cases (Gregor & Benbasat, 2006). However, due to the emergence of technologies associated with AI such as Machine Learning (ML), voice recognition, and natural language processing, new generations of SPAs have emerged such as Amazon’s *Echo*, Google’s *Google Assistant* and Apple’s *Siri*. They can now be applied in a wide range of use cases spanning from everyday tasks such as ordering consumer goods to more specialized tasks such as helping users track their expenses (Knote et al., 2019). There are various terms describing these assistants – for example conversational agent, chatbot, virtual

assistant, digital assistant -, who all are based on the idea of interacting with users via natural language (e.g., Rzepka & Berger, 2018). In this regard, in order to cover several types of systems, we refer to SPAs as AI-based systems embedded in personal technologies designed to assist users by interacting in a text or voice-based conversation (Feine et al., 2019).

From a sociotechnical perspective and compared to other entities of IS, the novelty of SPAs lies in two major aspects, which potentially fundamentally affect user perceptions (i.e., trust): the way in which SPAs *interact* with users as well as the degree of *intelligence* they employ thereby (Maedche et al., 2016). Thus, they manifest the characteristics of two distinct but interrelated system classes - interactive and intelligent IS: on the one hand, based on anthropomorphic features interactions with SPAs are increasingly moving towards the level of interpersonal communications (Pfeuffer, Benlian, et al., 2019), including the establishment of emotional bonds (Purington et al., 2017). At the same time, the pervasiveness of invasive technologies embedded in these systems as well as their autonomous nature raises questions of accountability and data security (Cowan et al., 2017). Moreover, the rising intelligence of SPAs comes with issues of interpretability of their behavior through users. This may explain why users still only reluctantly adapt and use these systems despite their potential (J. D. Lee & See, 2004).

B.2.2 Trust in Information Systems

One of the most important factors driving the adaption and use of complex and increasingly automated technical artifacts such as SPAs is trust. Traditionally, trust research in IS has been focusing on studying relationships among human beings that are mediated by an IS. However, due to developments such as increasing automation (J. D. Lee & See, 2004), IS have itself become an integral part of trust relationships in a wide area of usage scenarios. Automated systems are not only used to mediate trust relationships between human beings, but to support their users in achieving specific goals, thereby exhibiting agency on their behalf. Thus, these systems become themselves trustees in a trust relationship between the human user and a respective IS (Söllner, Hoffmann, et al., 2016). According to the trust definition of R. C. Mayer et al. (1995), users therefore need to exhibit willingness to be vulnerable to the actions of an autonomous IS “*based on the expectation that the other [i.e., SPA] will perform a particular action important to the trustor [i.e., user], irrespective of the ability to monitor or control that other party [i.e., SPA]*” (R. C. Mayer et al., 1995, p. 712).

In the past years, IS got increasingly interactive – specifically in regards to exhibiting anthropomorphic features – and intelligent – based on advancements in the domain of ML (Maedche et al., 2016), which is why trust in these systems may not anymore entirely be explainable with current insights. On the one hand, SPAs are able to act and interact in an increasingly human way. Thus, the boundary between man and machine becomes increasingly blurred from a user perspective, which has important implications for theory and practice. Especially, the suitability of the theoretical basis on which trust in the system is studied becomes a relevant question. On the other hand, AI-infused systems raise the opaqueness and complexity for the user (Maedche et al., 2019), therefore magnifying the issue of trust. It is argued that building trust is an essential means to address complexity and uncertainty because humans cannot have complete knowledge of most systems' inner processes. Additionally, as these systems continuously learn and adapt their behavior accordingly, there is an increased need to study trust in these systems from a longitudinal perspective (Holliday et al., 2016). In this work, we focus on SPAs as one concrete instantiation of AI-based IS, which we link to trust research as one combined unit of analysis based on a SLR.

B.3 Method

We conduct a SLR within the research field of trust in SPAs according to the principles and practices suggested by Webster & Watson (2002) and vom Brocke et al. (2009, 2015). Overall, the scope of the SLR can be structured along the dimensions of process, source, coverage, and techniques (vom Brocke et al., 2015): based on a *sequential search process* in four data bases, publications from IS literature and related fields such as business and human computer interaction as a *source* are identified. The literature search aims to reach a *representative coverage* of the distinct perspectives on the research field of trust in SPAs. Therefore, to establish the basis for the analysis and conceptualization, we used a *comprehensive* set of techniques (i.e., keyword search, backward search, and forward search). To reach a high level of reproducibility and transparency of our research, we describe in this section the single methodical steps that we undertook:

Selection of search strings: Aiming at covering literature that focuses on the combined unit of analysis of trust and SPAs, we select ("smart" OR "intelligent" OR "cognitive") AND ("assistant" OR "system") AND "trust" as the initial search string that we use as a starting point for the literature search process. The initial search string is constructed

rather broad taking into account a variety of key word permutations to not neglect relevant research. It needs to be noted here that we apply the search string considering the particular variations the single keywords can exhibit such as singular and plural, and the use of hyphens or no hyphens. As we seek for papers that conduct research with a focus on trust and SPAs as a combined unit of analysis, the single parts of the initial search string should appear in close proximity to each other in the papers. Thus, we choose to search in title, abstract, and keywords of publications.

Selection of databases: We apply a respective search in IS databases that contain a variety of IS journals and conferences to not restrict our search scope upfront and to cover more recent research as we assume that trust in SPAs is a young and emerging research field. Covering the latter aspect would not be assured by a journal-only-based literature search as processing journal reviews takes significantly longer than reviewing for conference proceedings. Consequently, we select five databases covering a wide range of Information Systems (IS) literature to assure a representative coverage of our literature search. The databases that we select are in particular the database of the Association for Information Systems (AISel), EBSCO, Sciencedirect, the database of the Association for Computing Machinery (ACM), and ProQuest.

Refinement of search strings: During our literature search process that starts with the search string described above, we iteratively refine and adapt the keywords to take into account the learnings from and our enhanced understanding of the field that emerges during our SLR. The final search string used in the SLR is ("smart" OR "intelligent" OR "cognitive" OR "conversational" OR "AI") AND ("assistant" OR "system" OR "agent" OR "application") AND "trust". It extends the initial search string by taking into account a larger set of synonyms used to describe SPAs and concepts related to the concept of “smartness” such as Artificial Intelligence (AI). The literature search conducted in title, abstract, and keywords of the publication texts, results in 1,168 hits, which still comprises duplicates and potentially irrelevant papers.

Selection of papers: During a first screening step, we focus on screening the title, abstract, keywords, and research domain of the papers and only consider papers that use English as their publication language. This process results in 45 papers, which undergo a detailed full-text screening that determines if a paper is finally considered relevant for deeper analysis. We label a paper as “relevant for further analysis” if it conducts research with a central focus on trust and SPAs as a combined unit of analysis. Consequently, during full-text screening, we remove literature that only marginally or trivially issues

the intersected unit of analysis “trust and SPAs”, such as for example paper that measured trust as one of many variables, but did not discuss this effect further. This leads to 35 relevant hits. After removing duplicates that stem from choosing a database-oriented literature search, 32 papers remain to be considered for further analyses. Table 11 provides an overview of the total hits and relevant search results structured along the particular databases to account for reproducibility and transparency of the SLR.

Table 11. Results of the literature search process

	Databases									
	AISEL		EBSCO		Science direct		ACM		ProQuest	
Search String	Hits	Relevant	Hits	Relevant	Hits	Relevant	Hits	Relevant	Hits	Relevant
("smart" OR "intelligent" OR "cognitive" OR "conversational" OR "AI") AND ("assistant" OR "system" OR "agent" OR "application") AND "trust"	79	8	142	4	480	4	82	14	385	5
Number papers selected for further analysis from 1168 screened papers	With Duplicates:		35	+	3 Forward Search				=	50
	Without Duplicates:		32		15 Backward Search					

To seize the benefits of a comprehensive set of search techniques, backward search (+15 additional papers) and forward search (+3 additional papers) are conducted on top of the literature retrieved from using the search string in the five databases above (vom Brocke et al., 2009, 2015; Webster & Watson, 2002). Finally, this results in an overall number of 50 papers that are analyzed and conceptualized in a detailed manner in this paper.

Paper Analysis and Conceptualization: We analyze the 50 papers identified to be relevant for this work from a concept-centric perspective. Thus, according to Webster & Watson (2002) a concept matrix is created based on the literature search results. Respectively, all papers are analyzed according to the focal concepts used to investigate the combined unit of analysis “trust in SPAs”, according to the applied research method, and according to the contributions reached for theory and practice. This endeavor intends to conceptualize the distinct central research streams that constitute the combined research field of trust in SPAs, thus providing an integrated view on the latter. We use an iterative process guided by cross-validation discussions between two researchers, in which we analyze and aggregate the distinct concepts identified in the retrieved literature to higher-order, more abstract concepts that are merged into particular meta-perspectives on the research field of trust in SPAs. By iteratively cross-validating the conceptual insights that are abductively created from literature we aim for reproducible, transparent, and valid research results. However, we have to acknowledge that conceptualizing literature always contains a certain level of non-erasable subjectivity.

B.4 Results

Figure 11. shows that the number of identified publications has been steeply growing during the last years. The youngest paper is from 2019 and the oldest paper from 1999, when initial interest rose in light of the first expert systems being used in organizational contexts. The majority of papers has been published within the last two years, which supports our initial assumption that trust in SPAs represents an emerging research field.

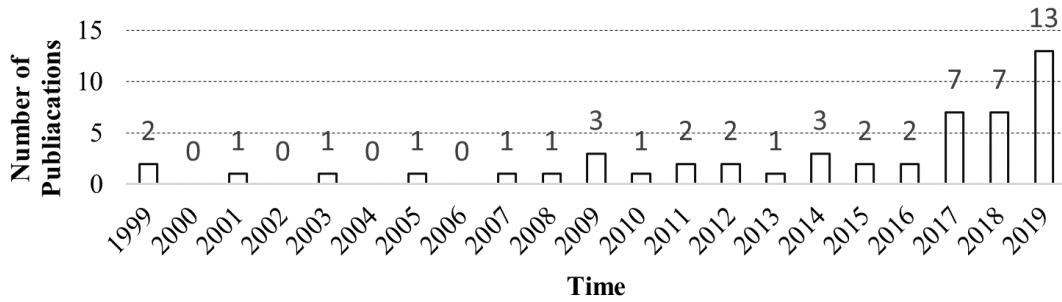


Figure 11. Number of publications over time

Based on a concept-centric analysis of identified papers, we were able to identify three main research perspectives, which constitute the research field of trust in SPAs: User Interface-Driven Trust (UIDT), Interaction-Driven Trust (IDT), and Explanation-Driven Trust (EDT). The research perspective with the most papers was IDT (26) followed by EDT (18) and UIDT (16).

Table 12. Definition and conceptual boundaries of the three perspectives on trust in SPAs

	Definition	Conceptual Boundaries
User Interface-Driven Trust (UIDT) in SPAs	UIDT in SPAs research deals with static design features of SPAs such as haptics and audio-visuals towards enhancing trust in the latter.	The scope of UIDT in SPAs research encompasses the effect of all components that make up the static appearance of SPAs on trust, including its audio, visual and haptic representation. This does not include the actions performed by the SPA.
Interaction-Driven Trust (IDT) in SPAs	IDT in SPAs research addresses the design of interactions between the user and the SPA on a timeline to form trust in the latter.	The scope of IDT in SPAs research encompasses the effect of the actual interaction behavior of SPAs with the user (and vice versa) on trust, including verbal and non-verbal behavior. This does not include the static representation of the SPA.

Explanation-Driven Trust (EDT) in SPAs	EDT in SPAs research focuses on creating trust in SPAs by varying the degrees of understandability and transparency provided to the user.	The scope of EDT in SPAs encompasses the effect of information and self-disclosure on trust in the SPA. This does not include the static representation of the SPA.
---	---	---

Overall, we strive for a precise and unambiguous description of the different research perspectives, in order to allow for a robust categorization of identified publications. Therefore, as Table 12 shows, we formulated concise definitions for the derived research streams and their respective conceptual boundaries. Furthermore, we highlight their interrelations according to the principles of construct clarity in order to achieve clear differentiations between the streams and to maintain internal consistency within the streams (Suddaby, 2010).

To make this conceptualization of trust in SPAs more graspable, we would like to discuss the different perspectives using the example of a Smart Home Assistant. Those agents can for example be used to obtain a wide range of information, to order products or control other smart home devices. UIDT addresses questions such as should the assistant speak with a male or female voice to the user. IDT, on the other hand, focuses on questions as to whether the assistant should communicate in a friendly or more emotionally neutral way. Finally, EDT is concerned with question as to whether and how the assistant should explain recommendations or actions takes such as automatically ordering a product.

In the following sub sections, we present the conceptually derived research streams that constitute the combined research field of trust in SPAs by elaborating on the particular research perspectives and sub research streams of the latter.

B.4.1 User Interface-Driven Trust in Smart Personal Assistants

One sub-research stream deals with the effect of visual design features, which refers to interface-transmitted cues that can be perceived visually (Feine et al., 2019). A majority of studies within this sub-stream focuses on the effect of SPA embodiment on user trust (Jensen et al., 2010; Krogue, 2017). For instance, it was found that humans perceive trust relationships with humans differently than with avatars such that humans are better in predicting the trustworthiness of humans than of avatars. However, the trustworthiness learning rate is similar, whether interacting with humans or avatars (Riedl et al., 2014). In regards to the effect of embodiment, there are somehow mixed results. In a

study conducted with children, they rated the most visually embodied character as the most trustworthy SPA in a game scenario (Druga et al., 2017). Moreover, it was found in another study that embodied SPAs are associated with greater trust resilience, a higher resistance to breakdowns in trust, and that these effects were magnified by greater uncertainty. However, once the different SPAs incorporated human-like trust repair behavior the effect was largely erased (de Visser et al., 2016). Moreover, in a survey ranking different trust mechanisms it was found that visual appearance is the least important for users of robo-advisory services (Mesbah et al., 2019). These findings suggest that the effect of appearance is highly dependent on the context. This is supported by another study showing that gender fit between the avatar and the user may present an important antecedent for trust formation (Nunamaker et al., 2011). Another study reports that male avatars are being experienced as more trustworthy as female avatar in interview scenarios (Nunamaker et al., 2011). Finally, authors investigated the effects of demeanor and could for instance show that a smiling SPA increased trustworthiness (Elkins & Derrick, 2013).

Another sub research stream in the dimension of UIDT in SPAs is related to the effect of auditory design features on user trust, which refer to cues that can be perceived audibly (Feine et al., 2019). Most authors, studied the effects of response modality, such as the questions if and which kind of voice is perceived as more trustworthy compared to using a chat interfaces (Gregor & Benbasat, 2006; vom Brocke et al., 2009, 2015). Generally, identified studies show that human-(like) voice carries important cues that evoke perceptions of social presence resulting in higher levels of perceived trustworthiness. Thereby, there are a number of SPA speech properties such as pitch contour and flanging increments that classify speech along a machine-to-human spectrum (Muralidharan et al., 2014). Along this spectrum, several authors found that human speech had higher ratings of trust than machine-like speech, which also translated into higher compliance with the SPAs recommendations (Qiu & Benbasat, 2009; Yu et al., 2019). Text-to-Speech (TTS) voice leads to lower levels of social presence and was perceived as less trustworthy than text. (Qiu & Benbasat, 2009). Moreover, authors found that expression modality seems only to be relevant for initial trust formation, since it was found that vocal pitch only influenced trust perception early in the interaction process (Elkins & Derrick, 2013).

B.4.2 Interaction-Driven Trust in Smart Personal Assistants

The identified research stream of *Interaction-Driven Trust (IDT) in SPAs* addresses the design of interactions between the user and the SPA on a timeline to form trust in the latter. Thereby, the authors generally discuss different verbal and non-verbal features that make up the behavior of the SPA employing a processual perspective:

The first sub research stream (i.e., verbal features) can be divided into contributions that focus on content and on conversational style (Feine et al., 2019). Thereby, we found that most publications address the latter. In regards to content, initial work identified the use of small talk in the SPAs interaction with the user as an effective strategy for building trust (Gregor & Benbasat, 2006). Moreover, being critical to the users wishes was mentioned as another trust-building strategy (Vaccaro et al., 2018). Studies addressing conversational styles in general found a positive effect of relational strategies. Thus, socially oriented SPAs interjecting an informal and friendly conversational style lead to enhanced perceptions of interactivity and trust in the system (Chattaraman et al., 2019). In comparison, task-oriented SPAs without any deliberate social-emotional capabilities were trusted less. This effect seems to remain stable over time, since even after weeks of interacting with a SPA, relational agents were perceived as significantly more trustworthy than task-oriented agents (T. W. Bickmore & Picard, 2005). Thereby, SPA responsiveness was found to be a major antecedent (Wong-Villacres et al., 2019). One study highlights the importance for SPAs to learn from human-human interactions on how to build trust, but recommends to not just mimic it. Instead, human-SPA conversations may need to be treated as a new genre of interaction as trustworthiness was discussed exclusively in utilitarian terms by interviewees in the same study – responses related to security, privacy, and transparency over emotional trust (L. Clark et al., 2019).

The second sub research stream (i.e., non-verbal features) investigated primarily the effect of the degree of autonomy (i.e., proactivity) exhibited by the SPA. Thus, it was shown that autonomy may lead to a more human-like appearance evoking a feeling of social presence and therefore inducing trust (J. G. Lee et al., 2015). However, another study suggests that there is a need for a fit between the SPA's autonomy and user preferences to maintain trust. In a smart home environment it was shown that the users level of comfort with the SPA's level of autonomy was depended on personality types and task characteristics (Hammer et al., 2015). To give users a perception of control different mechanisms are proposed – personalization-driven control, task-driven control, and mechanism that allow direct control (Cesta & D'aloisi, 1999). Specifically, providing

users with different alternatives has proven to provide users with a sense of control, which increases trust in the SPA (Cummings et al., 2010).

B.4.3 Explanation-Driven Trust in Smart Personal Assistants

Explanation-Driven Trust (EDT) in SPAs refers to varying the degrees of understandability and transparency provided to the user. Within the analyzed papers, we could identify two sets of strategies that were applied to increase transparency and understandability – explanations and interactive machine learning (Meza Martínez et al., 2019).

In the first sub-research stream, authors investigate the effect of different types of explanations for increasing trust in SPAs. Generally self-disclosure was deemed an effective measure to signal trustworthiness to users (Cesta & D’aloisi, 1999; Cummings et al., 2010; Hammer et al., 2015; J. G. Lee et al., 2015). In this regard, it was suggested that the SPA should frequently communicate what kind of data it needs to generate value for the user (Cesta & D’aloisi, 1999). Apart from one study reporting that only users with low task familiarity are susceptible for explanations provided by SPAs (Schaffer et al., 2019), most of the studies see generally a positive effect in respect to trust formation (Cummings et al., 2010) by raising transparency and focusing the users focus on system ability (Holliday et al., 2016). Even providing placebic explanations was found to raise trust in the SPA (Eiband et al., 2019). However, informative explanations were more effective in building trust. Thereby, it was shown that even laypersons are able to understand the basic logic behind ML-models. In this regard, rule-based and keyword-based ML-models ranked high in understandability, while similarity-based ML models were harder to grasp for users resulting in less trust into the actions of the SPA (Stumpf et al., 2009). An often-addressed topic is the analysis of explanation types for improving the intelligibility of SPAs. In this regard, it was proposed that explanations based on justifications following a structured argumentation approach and addressing the reasoning of the system’s behavior would evoke higher perceptions of trustworthiness (T. W. Bickmore & Picard, 2005; Chattaraman et al., 2019). Finally, the effect of explanations on user trust was investigated from a longitudinal perspective. It was shown that explanations increase trust levels in the short term, but have no effect in the long term. However, without any explanation shown trust levels in regards to the SPA degenerated steadily (Holliday et al., 2016).

The second sub-research stream, smaller in numbers than the previous one, is concerned with the ability of SPAs to increase system transparency by enabling users to influence its recommendations (i.e., interactive ML) (Meza Martínez et al., 2019). Thus, users

were given the opportunity to set constraint thresholds or to change algorithm weights (Cummings et al., 2010). Thereby, it was shown that providing the user with the opportunity to test the SPA is one of the most effective means to increase trust (Mesbah et al., 2019).

ement items do not represent a complete, exhaustive representation of the respective constructs, but instead are indicative of a certain construct being present in the organization of a responding participant.

B.5 Discussion and Development of a Research Agenda

In this section, we aim to discuss the contributions of our SLR and propose a preliminary research agenda that provides first promising points for future research on trust in SPAs and illustrates how they can be positioned based on our conceptualization.

As our SLR shows, the three identified research streams enable a distinct perspective on studying trust in SPAs, which also relates to the theoretical lenses that are applied. While UIDT and IDT are linked primarily to enhance trust by creating a sense of social presence, EDT mainly relies on perceived transparency. Additionally, each of the research streams can be distinguished by its time perspective. While UIDT is important to build initial trust as the interface represents the first point of interaction, IDT and EDT additionally imply a longitudinal perspective. However, this conceptualization only serves as a starting point for the elaboration of the different theoretical lenses, which can be applied to study trust in SPAs. This is important to allow researchers to apply a more nuanced perspective when studying trust in SPAs.

Building on the insights that have been gained through this SLR and by linking our conceptualization to our theoretical background, we propose the following preliminary research agenda as presented in Table 13. Thereby, both possible questions that arise from the research streams and some overarching research questions are addressed:

Table 13. Preliminary research agenda on user trust in SPAs

	Research Opportunities	Corresponding Research Questions
UIDT	Haptics are associated with user perceptions (i.e., trust) (Feine et al., 2019); however, this effect has not yet been sufficiently addressed in the context of SPAs.	How do haptics (e.g., temperature, tactile touch) influence trust?
	Interfaces are usually characterized by multiple features; however, the effect of specific feature combinations has not yet been adequately explored.	How do specific feature combinations (i.e., based on different system archetypes; Knote et al., 2019) compare in regards to user trust?
IDT	As our SLR has shown, content may significantly influence trust, but still represents an area that has not yet been sufficiently addressed.	How do different content features (i.e., praise; Feine et al., 2019) of a conversation affect trust?
EDT	As new types of explanations emerge based on technologic advances (Rzepka & Berger, 2018), there is a need to study their efficiency in regards to creating trust.	How do different types and instantiations of explanations (i.e., attribution-, example- or model-based explanations affect trust?
	Although interactive ML has been identified as a major source of trust (Mesbah et al., 2019), there are still few insights on the effect of different types of interactive ML (Meza Martínez et al., 2019) on trust.	How do different types of interactive ML affect trust?
Overarching	There are few insights on the relative importance of the identified types of trust (e.g., our conceptualized research streams).	How do different types of trust (e.g., UIDT, IDT, EDT) compare in regards to their effectiveness for building trust?
	Different applications of SPAs may have different presuppositions for trust formation (Pfeuffer, Benlian, et al., 2019); however, the impact of context remains under-researched.	How do contextual factors (e.g., user group, time) influence the effect of different SPA elements and characteristics on trust?
	There is a lack of design-oriented studies (Pfeuffer, Benlian, et al., 2019) and successful SPA designs that foster trust.	How to leverage theory to design more trustworthy SPAs?

All in all, eight research questions emerged both on the level of the individual research streams as well as some overarching research questions, which bears to the relevance of

the topic of trust in SPAs. Across all research streams, we identified research opportunities in regards to the influence of specific features. Moreover, as the use context of SPAs may entail different presuppositions for trust formation, especially based on their high degree of adaptivity (Maedche et al., 2019), we propose to increasingly investigate the influence of contextual factors. Especially, the influence of time has not yet been sufficiently addressed. A good starting point could be to analyze which design features are important for building initial trust in SPAs and which are more important to uphold trust in the long-term. Additionally, we need to create insights across different domains on how to design SPAs to increase their trustworthiness (e.g., Pfeuffer, Benlian, et al., 2019). Finally, from a methodological perspective, as most of the identified studies are based on laboratory experiments, we recommend to increasingly use field experiments to ensure external validity and, thus, to be able to provide stronger insights for practitioners.

In sum, these emerging agenda points may serve as a first foundation for studying trust in SPAs.

B.6 Limitations

Although we attempted to analyze the identified literature on trust in SPAs as rigorously as possible, there are a number of limitations to this SLR. First, of course the scope of our SLR is not fully exhaustive. However, in order to reach representativity, we chose to conduct a database-oriented search instead of a journal-based search. This enabled us to also consider conference proceedings, which include recent publications, which are especially important when the analyzed research field is still young and emerging such as research on trust in SPAs. We restricted the keyword search to title, abstract, and keywords, since we aimed for identifying publications where the keywords appeared in close proximity to each other for the combined unit of analysis was trust in SPAs. Therefore, it can be argued that the initial screening is limited in scope, but we thoroughly analyzed identified publications based on a concept-centric approach. Moreover, we conducted a rigorous back and forward search. Finally, we did not consider the downstream effects of trust in SPAs (e.g., possible negative effects of too much trust; Benlian et al., 2020), which may represent an interesting research avenue for future literature reviews in this area.

B.7 Conclusion

In this paper, we conducted a Systematic Literature Review (SLR) on trust in Smart Personal Assistants (SPA). SPAs bear an immense potential for economic and societal impact, since they fundamentally change the way how individuals perform tasks, use services and interact with organizations. However, they are only hesitantly adopted by users amongst others due to a lack of trust. Building on a concept-centric analysis of 50 publications, we derived three main research perspectives, which constitute the research field of trust in SPAs: User Interface-Driven Trust (UIDT), Interaction-Driven Trust (IDT), and Explanation-Driven Trust (EDT). UIDT deals with static design features of SPAs such as haptics and audio-visuals towards enhancing trust in the latter. IDT refers to trust that is elicited by the design of events between the user and the SPA on a timeline. While UIDT and IDT in SPAs focus on creating a sense of social presence, which has been identified as an important antecedent of trust, EDT in SPAs aims at creating a sense of transparency by varying the degrees of understandability and transparency provided to the user. Based on the results of our analysis, we propose a preliminary research agenda to spark and guide future research in this nascent field. We intend to contribute to the body of knowledge of trust in Artificial Intelligence (AI)-based systems by proposing an integrated theoretical framework and associated research agenda for studying trust in SPAs. This framework may serve as a starting point for the elaboration on different theoretical lenses which can be applied to study trust in SPAs from a more nuanced perspective. Ultimately, this may also enable practitioners to build more trustworthy SPAs as we introduce new terminology that facilitates the sharing of design knowledge on the effects of different features on trust.

Paper C – Gendered Bots as Surrogates of Trust? How Voice Gender and Communication Style Affect Interface Trust and Firm Perceptions

Table 14. Bibliographical metadata Paper C

Title	Gendered Bots as Surrogates of Trust? How Voice Gender and Communication Style Affect Interface Trust and Firm Perceptions
Authors	Zierau, Naim
Publication outlet	<i>Proceedings of the 56th Hawaii International Conference on System Sciences</i>
Publication type	Conference Paper
Publication year	-
Rating (Jourqual 3)	C
Publication status	Submitted

Abstract

Many firms have implemented female voices in voice user interfaces (VUIs) to increase customer acceptance. Meanwhile, little is known about the conditions under which voice gender affects interface trust and firm perceptions. Therefore, we designed an online experiment to investigate a novel boundary condition to the effect of voice gender on trust manifested in the conversational design. We find that a male voice paired with a social-oriented conversational style reduces interface trust. This can result in less favorable firm perceptions (in terms of the service experience and benevolence attributions). For a task-oriented conversational style, we do not find an effect of voice gender on interface trust. The finding of this research highlight how fundamental insights into interpersonal communication can be harnessed to create more trustworthy user experiences with VUIs in service contexts. Further, they have important practical implications for the theory-driven conversational design of VUIs.

Keywords

Voice User Interface, Voice Gender, Conversational Style, User Trust, Firm Perceptions

C.1 Introduction

Voice user interfaces (VUIs) transform how users interact with firms along the increasingly technology-mediated customer journey (Diederich et al., 2020; Seaborn et al., 2021). They enable firms to automate customer engagement with users still experiencing the dynamic of human interaction, almost as if they were talking to a service employee (Huang & Rust, 2021). For example, nowadays, customers can conduct money transfers by speaking to Bank of America's voice assistant Erica (Fuscaldo, 2019) or receive an insurance quote using the Alexa skill of Liberty Mutual Insurance (Mauze et al., 2018). Increasingly, VUIs enable customers to conduct complete automated service requests more intuitively and naturally than using traditional technologies (Wirtz et al., 2018). Indeed, recent industry reports indicate that the number of VUIs is rising exponentially (e.g., 8.4 billion VUIs by 2024, Laricchia, 2022b). Likewise, customer interest is picking up: 72% of US customers have used a VUI at least once (P. Hayes & Wagner, 2018), which indicates the potential of VUIs for service provision.

Despite customers' interest in the new service channel, the adoption of VUIs is still relatively low (Lovett, 2019). Specifically, a lack of trust is cited as a barrier to the widespread adoption of VUIs (de Bellis & Venkataramani Johar, 2020). In this context, the emergent voice-based interaction paradigm of VUIs could play a key role (Diederich et al., 2022). VUIs mimic human communication in spoken form, which allows for an interactive dialogue and the use of para-lingual cues (Moffett et al., 2021). These capabilities may fundamentally affect how users develop trust in these information systems (Rockmann & Northcraft, 2008).

An essential element of voice-based interactions that has been linked to trust is the voice itself (O'Connor & Barclay, 2017). Voice carries implicit para-lingual cues (i.e., voice pitch) that allow listeners to make inferences about a person's personality traits (Reinkemeier & Gnewuch, 2022; Sutton et al., 2019). Interestingly, most contemporary VUIs utilize female voices (as the default option) (Hefti, 2022). According to a recent study, up to 77% of all virtual assistants embody features that can be classified as female (Feine et al., 2020) – among them leading voice technology platforms from firms such as Amazon ("*Alexa*"), Apple ("*Siri*"), and Microsoft ("*Cortana*"). This burgeoning trend raises the question of how voice gender affects trust in the VUI in customer service settings with which downstream effects on users and firms.

From the literature on social heuristics, we know that there are gendered preconceptions about which voice sounds the most friendly and helpful (McAleer et al., 2014). Specifically, research shows that generally, humans rate female voices more trustworthy than male voices (McAleer et al., 2014; O'Connor & Barclay, 2017). In human-computer interaction, the findings indicate that trust attributions based on voice gender are highly context-dependent (e.g., on the type of task, Nass et al., 1997). Building on this research, this paper aims to address two gaps. First, although we know that people have gendered preconceptions about how a person communicates (Fiske et al., 2002), no research investigates how conversational design itself impacts the effect of voice gender on interface trust. Specifically, the VUI's conversational style is a much-used conversational element that, when combined with voice gender, could have a unique impact on the user experience (Hannah & Murachver, 1999). Second, there is a lack of evidence on the downstream impact of gendered VUIs on the firms they represent. Hence, the main objective of this paper is to examine how voice gender and conversational style influence trust in the interface and firm perceptions. Specifically, we aim to answer the following research question:

RQ: *How do voice gender and conversational style influence interface trust and firm perceptions in VUI-mediated service encounters?*

To answer this research question, we recruited 211 participants from Prolific and randomly assigned them to a 2 (voice gender: female vs. male) x 2 (conversational style: social-oriented vs. task-oriented) between-subject online experiment. Specifically, participants were tasked to file an insurance claim with a VUI. Our results show that there is no direct effect of voice gender on trust when the interface used a task-oriented conversation style. However, we found that a male voice paired with a social-oriented conversational style reduces interface trust, resulting in a more negatively-valenced service experience and lower attributions of benevolences towards the firm. Thus, we contribute to the literature on trust in automated service interfaces by illuminating a novel boundary condition to the effect of voice gender on trust in the VUI. Moreover, we systematically investigate downstream consequences of trust in VUIs on firm perceptions within a fully conversational service interaction. These findings have important implications for the effective design of VUIs in service contexts.

In what follows, we first provide an overview of the theoretical background and related work on trust in automated systems, trust in VUIs, and voice gender. Subsequently, we develop a set of hypotheses on how voice gender and conversational style together alter trust in the interface and firm perceptions. We then present the results of an experimental

study designed to test our theorizing. Finally, we discuss the theoretical and practical implications for the conversational design of VUIs and derive avenues for future work.

C.1.1 Trust in Automated Interface

Trust is an integral part of social interactions, which fundamentally determines how people behave towards other people and organizations (R. C. Mayer et al., 1995; Rousseau et al., 1998). Moreover, with increasing automation, it can be argued that IS themselves are becoming recipients of trust (J. D. Lee & See, 2004). As such, in an increasing number of service encounters, users need to rely on an automatic interfaces to solve their requests (Söllner, Hoffmann, et al., 2016). In the context of this study, we define trust as “*a psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or behavior of another [i.e., the VIU]*” (Rousseau et al., 1998, p. 395). Moreover, according to Söllner, Hoffmann, et al. (2016) different entities that are relevant to a context have to be considered. Specifically, in the customer service context, trust toward the service interface may interact with trust toward the service provider. Thus, trust can transfer from one entity to another (Söllner, Hoffmann, et al., 2016).

By this definition, the voice of a VUI may affect interface trust by triggering specific social heuristics that influence a user's expectation about the VUI's intentions. Moreover, the design of a VUI can both affect interface and firm perceptions.

C.1.2 Voice User Interfaces and Trust

The growing proliferation of voice-based technology platforms is gradually shifting service operations toward VUIs (Seaborn et al., 2021). VUIs are powered by recent computational advances that combine speech recognition, natural language processing and text-to-speech generation to enable a fully voice-mediated service interaction (Seaborn et al., 2021). VUIs promise to improve the service experience based on a more personalized and human-like interaction experience (Diederich et al., 2022; Larivière et al., 2017). However, many users are skeptical about this new form of human-computer interaction (Sezgin et al., 2020). In particular, a lack of trust is seen as an obstacle to their further dissemination across the customer journey (de Bellis & Venkataramani Johar, 2020; Pfeuffer, Benlian, et al., 2019). Specifically, many firms fear reputational damage from the misguided design of VUIs (Kaneshige & Hong, 2018).

Hence, developing more trustworthy interface designs is critical to driving the adoption of VUIs (Thiebes et al., 2021). While there is already a plethora of IS research on trust

in various technology platforms (Söllner, Benbasat, et al., 2016), further research on trust in VUIs is warranted based on their unique characteristics. First, the interaction flow follows a dynamic, voice-mediated interaction protocol in which the user and the interface communicate alternately by taking turns (Diederich et al., 2022). The dialogic form of interaction has also expanded the use of language towards more informal forms of communication. For instance, VUIs use trivial acknowledgments to respond to users (Thomaz et al., 2020). Second, VUIs can use para-lingual cues such as pitch, speed, or volume to convey emotions and personality (Hildebrand et al., 2020). All this allows VUIs to interact with users in what was previously considered a hallmark of human communication: voice-based dialogue (Jurafsky & Martin, 2017). This capacity, however, may fundamentally affect user experiences (Diederich et al., 2022). Specifically, there is a need to better understand how VUIs as a new interaction paradigm affect trust perceptions in the interface and the firms they represent (Diederich et al., 2022).

C.1.3 Related Work on Trust in Voice Gender

One feature that can uniquely influence trust in voice-mediated interactions is the voice itself (McAleer et al., 2014). Accordingly, the way a person speaks can activate social heuristics based on which another person judges trustworthiness (Eyssel & Hegel, 2012). Gender is such a key trigger of trait inferences based on social heuristics (Schniter & Shields, 2020). For example, women are more likely to be perceived as friendly and men as capable (Fiske et al., 2002). According to social response theory, people are likely to attribute these human-like characteristics to non-human entities such as an IS (Nass & Moon, 2000). For example, studies have shown that imbuing a robot with a voice alone led users to label them as female or male and assign specific task roles to them based on "their" gender (Tay et al., 2014). Given that many VUIs use female voices, this raises the question of how and when the gender of the voice affects trust in VUIs and the relationship with the firms employing the VUI.

In Table 1, we summarize and compare prior work on gendered VUIs. Specifically, three key factors help to understand the unique insights of this work: (1) the operationalization of the interface, (2) the type of examined boundary conditions, (3) and whether outcomes focused more on the interface as opposed to firm perceptions. First, most prior works examined the effect of gendered VUIs by operationalizing voice as an output, i.e., how an interface responds. For example, Tay et al. (2014) examined how different synthetic voices affected trust towards a robot in different task contexts when played to

participants. However, few studies examined a fully conversational interaction that requires a dynamic dialogue between the user and the VUI to conduct a service task (see Tolmeijer et al., 2021). Second, previous research focused on studying boundary conditions related to task characteristics (e.g., whether the task can be considered feminine or masculine; see Nass et al., 1997; Tay et al., 2014) or user characteristics (e.g., the effect of gender congruity between user and interface; see Reinkemeier & Gnewuch, 2022). We are aware of no prior work that examines the unique characteristics of the conversational design during an interaction, such as the specific conversational style employed by the interface. Third, the outcomes studied in earlier work focused predominantly on interface perceptions (such as stereotypical trait attribution; see Nunamaker et al., 2011; S. Song et al., 2020; Wong et al., 2019) as opposed to evaluating the firm deploying the interface.

Table 15. Review of Relevant Literature on Voice Gender

Source	Study	Main Results
Nass et al., (1997)	I: Output only / Static B: Task Type O: Interface Perceptions, User Behavior	Different voice genders were perceived as more favorable depending on the task context.
Nunamaker et al. (2011)	I: Output only / Static B: None O: Interface Perceptions	Voice gender led to stereotypical trait attribution. Specifically, the “male” interface was perceived as being more trustworthy.
Tay et al. (2014)	I: Output only / Static B: Task Type O: Interface Perceptionss	Perceived task-gender congruity increases interface trust.
Wong et al. (2019)	I: Input and Output / Conversational B: Task Type O: Interface Perceptions	Interface–user gender congruity increases interface trust.
A. Kim et al. (2019)	I: Output only / Static B: Relationship Type O: Interface Perceptions	No significant effect of voice gender on trait attribution (competence, warmth)
S. Song et al. (2020)	I: Output only / Static B: Voice Pitch	No main effect of voice gender on interface trust and reliance. Significant interaction effect between pitch, user gender, and voice gender: male users perceived

	O: Interface Perceptions, User Behavior	high-pitched male voices as less trustworthy.
Tolmeijer et al. (2021)	I: Input and Output / Conversational B: Task Type O: Interface Perceptions	Female voices were trusted in an assistance task, male voices more in a compliance task.
Reinkemeier & Gnewuch (2022)	I: Input and Output / Scenario-Based B: Agent Personality O: Interface Perceptions	Interface-user gender congruity did not affect interface trust. Matching personalities of interface and user increased trust.
This Study	I: Input and Output / Conversational B: Conversational Style O: Interface perceptions, Firm perceptions	Significant interaction between conversational style and voice gender: Male voices are trusted less when using a social-oriented conversational style. Higher interface trust enhances delivery experience and benevolence attributions towards the firm.
<i>I: Interface Operationalization; B Boundary Conditions, O: Investigated Outcomes</i>		

As Table 1. shows, the current work investigates the effect of voice gender in a fully conversational service experience, examines a critical boundary condition (which conversational style the interface employs), and examines both perceptual outcomes relevant for users and firms.

C.2 Hypothesis Development

To illustrate how VUIs will affect interface trust as well as service and firm perceptions, we utilize the research framework shown in Figure 12.:

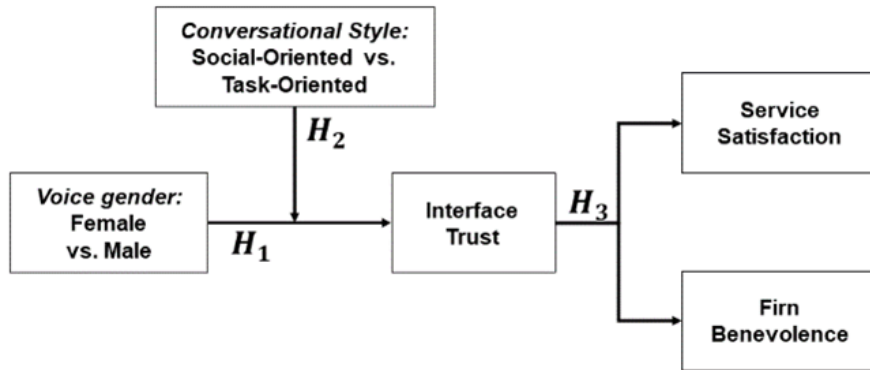


Figure 12. Research Framework

According to social response theory (Moon, 2000), humans tend to assign human characteristics to computer interfaces based on social heuristics they learn in everyday life. These heuristics are often inspired by social categories related to salient features of a person (e.g., gender) (Crisp & Hewstone, 2007). These heuristics are prone to cognitive biases following societal stereotypes (Eyssel & Hegel, 2012). One primary target of these stereotypes has been gender. According to the stereotype content model, these stereotypes boil down to the following dimensions: women are usually seen as warm and friendly, while men, on the other hand, are seen as competent and capable (Fiske et al., 2002). Prior research has shown that humans use the same stereotypes to evaluate the suitability of a computer interface for a specific task when using a gendered voice (Nass et al., 1997). In the customer service context, it is often assumed that companies act purely in their self-interest (Monti et al., 2014). Against this background, we assume that the qualities attributed to women, such as warmth and friendliness, should reinforce a perception of goodwill and thus increase trust in a service interface. Hence, we posit:

H1: *Interacting with a VUI with a female instead of a male voice increases trust in the interface.*

According to component fit theory, the perceived (mis-)match between different design elements of a computer interface can uniquely influence user experiences (Isbister & Nass, 2000). The level of consistency between design elements affects user experiences beyond their individual-level effects. Specifically, the level of consistency has been shown to determine fluency in human-computer interaction (Isbister & Nass, 2000) and thus to affect trust in the interface (Qiu & Benbasat, 2009). It can be assumed that especially a mismatch is salient for the user and has an impact on the interface experience.

Hence, following the stereotype content model (Fiske et al., 2002), we assume that pairing a social-oriented as opposed to a task-oriented conversational style with a gendered VUI should decrease trust in a VUI with a male voice. At the same time, following social response theory, we expect that using a task-oriented conversational style reduces the effect of voice gender on interface trust overall. Specifically, using a social-oriented rather than task-oriented conversational style has been shown to trigger enhanced levels of anthropomorphism (Feine et al., 2019). Consequently, we assume that using a task-oriented conversational style should decrease the tendency to apply gender stereotypes toward VUIs and thus decrease the effect of voice gender on trust. Taken together, we formally hypothesize:

H2: *A social-oriented (vs. task-oriented) conversational style decreases trust in the interface for a VUI with a male voice.*

Prior work shows that trust in automated service interfaces can positively affect numerous outcomes (Söllner, Benbasat, et al., 2016). In the context of this study, we assume that trust in the VUI positively affects firm perceptions in two significant ways. First, on the level of service delivery, satisfaction is conceptualized as a state where customer expectations are met or even exceeded (McKinney et al., 2002). As argued before, a lack of trust is critical for customers adopting VUIs to solve service requests (de Bellis & Venkataramani Johar, 2020). Hence, we assume that if a VUI can provide a trustworthy user experience, it should create an overall satisfactory service experience. Second, we expect trust in the VUI to lead to subsequent changes in firm evaluation. Prior research has shown that developing trust toward an entity has shown to solidify assumptions that this entity will act in one's best interest in future interactions (J. Cho, 2006). As such, we expect greater perceptions of trust toward the VUI, in turn, to lead to subsequent attributions of benevolence toward the firm employing the VUI (i.e., the assumption that a firm will act in one's best interest; R. C. Mayer et al., 1995). Taken together, we expect that:

H3: *Greater levels of trust in a VUI will lead (1) to a more positively-valenced service experience (2) and attributions of benevolence toward the firm employing the VUI.*

C.3 Research Method

C.3.1 Participants

We recruited a total of 215 participants from Prolific Academic. Each participant received a fixed compensation for participating in the study. All participants spoke English as a first language. 4 were excluded based on attention checks, resulting in a final sample of 211 participants ($M_{\text{Age}}=38.28$, $SD_{\text{Age}}=14.02$, 51.18% males).

C.3.2 Experimental Context

As a study context, we chose the insurance industry and specifically the task of filing an insurance claim for two reasons. First, the insurance industry represents a context of high economic and social importance. Over the past years, the claim filing process has been heavily standardized and automated to increase the efficiency of the claim management process. However, this growing automation has led to reduced customer satisfaction and a reported lack of “human interaction” (Blake, 2018). In recent years, the number of customer calls has risen sharply. Hence, firms are now aiming to deploy VUIs as a novel and more human-like form of service provision. Second, the insurance industry represents a context that is particularly sensitive to trust. Specifically, customers need to trust that insurance firms will act in their interest and that they will process an insurance claim promptly and benevolently (Guiso, 2021)

C.3.3 Experimental Procedure

The web experiment followed a fully randomized 2 (voice gender: male vs. female voice) x 2 (conversational style: transactional vs. relational) between-subjects design. In total, the study consisted of four individual phases: (1) randomization, (2) scenario description, (3) insurance claim, and (4) posttest. In the beginning, we provided the participants with a description of the experimental procedure and the scenario. In this vignette (Aguinis & Bradley, 2014), the participants act as customers of a fictitious insurance company called TER Insurance. Specifically, we asked the participants to file an insurance claim. Before the task, we presented the participants with a common claim-related scenario that described the events of a car accident (i.e., sequence of events, time, damage). Then, we asked the participants to report all information on the incident to a VUI. Across all conditions, the conversation flow was identical (i.e., the sequence and content of 6 conversational turns remained constant). We asked participants to either use a laptop or desktop computer (i.e., a non-mobile device) to avoid device-related

confounds. Immediately after completing the service interaction, we assessed the user’s perception of the service task using a questionnaire.

To conduct the proposed study and ensure maximum experimental control, we used a custom-made VUI building on Python and Google WaveNet. The design of the VUI was kept simple (i.e., not including any visual cues such as avatars that could also influence user perceptions; see also Figure 2.) to control for any factors that could have spurious effects on our key measures of interest. In addition, the dialogue flow was predefined to ensure that the order of questions was identical across all conditions. To produce the different voices, we used Google’s state-of-the-art text-to-speech generator WaveNet (Oord et al., 2016). As such, the participants went through a fully voice-mediated service interaction with individual prompts to complete the service task.

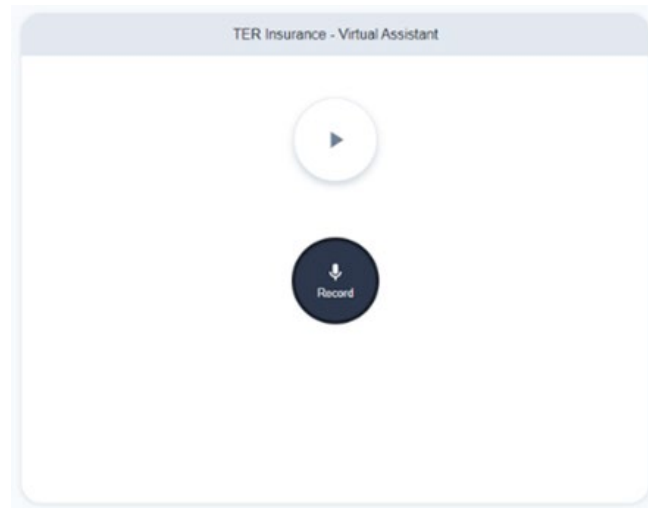


Figure 13. Experimental Platform

C.3.4 Experimental Stimuli

This study uses voice gender and conversational style as experimental conditions in a 2x2 factorial design.

To manipulate the voice gender, we differentiate between a male and a female voice. We conducted a pretest with 60 participants ($M_{\text{Age}}=33.62$, $SD_{\text{Age}}=12.71$, 48.33% males) on Prolific to test whether the female voice was in fact considered feminine and vice versa. The participants were randomly assigned to two conditions (voice gender: female vs. male). First, the participants listened to a recording of a random Harvard sentence in either the female or the male voice (Zhang, 2015). These are phonetic balanced sentences without a specific meaning. We then asked participants to rate the gender of the agent on a differential scale 1: masculine to 5: feminine. An independent t-test shows

that the manipulation was effective, as the female voice was rated significantly more feminine ($M_{\text{Male}} = 1.38$, $M_{\text{Female}} = 4.21$, $t = 17.272$, $p < .001$).

Regarding the manipulation of the conversational style, based on prior work in linguistics and communication research (Chattaraman et al., 2019), we altered the level of formality. For example, the transactional condition used a more formal style (e.g., “Provide an approximate estimate of the damage that has occurred”), and the relational condition used a less formal style (e.g., Could you please tell us how high the damage is?). We kept the number of words per conversational turn constant to control for any effects based on the length of the query and isolate the interaction’s conversational style ($M_{\text{TaskOriented}} = 16.00$, $M_{\text{SocialOriented}} = 16.16$, $t = .035$, $p = .97$). As a manipulation check, we conducted a pretest with 40 participants ($M_{\text{Age}}=29.58$, $SD_{\text{Age}}=9.34$, 50.00% males). The participants were randomly assigned to a two-cell between-subject experiment (conversational style: task-oriented vs. social-oriented). In the pretest, we used a text-based version of the interface to control for any spurious effects that may arise from the interaction of the conversational style and the voice itself. Following (Chattaraman et al., 2019) we tested the effectiveness of the conversational style manipulation based on the API engaging personality measure (scale adapted from Chattaraman et al., 2019; sample item: “The virtual assistant was friendly.”; 7 point-Likert scale, from 1: “Strongly disagree” to 7: “Strongly agree”; $\alpha_{\text{ConversationalStyle}} = .93$). The results from an independent t-test show that the manipulation was effective as the interface was perceived as more engaging when the interaction style was social-oriented instead of task-oriented ($M_{\text{TaskOriented}} = 2.43$, $M_{\text{SocialOriented}} = 3.72$, $t = 3.491$, $p < .01$).

C.3.5 Measurements

After going through the service interaction, we assessed users’ perceptions of interface trust (Srivastava & Chandra, 2018), service satisfaction (Bettencourt, 1997), and firm benevolence (Schlosser et al., 2006) on 7-point Likert scales, ranging from 1:strongly disagree to 7:strongly agree (see Table 2 for all items). We measured each construct by taking the mean of participants’ statements across those items. We used different methods to examine construct validity and reliability. First, all Cronbach’s alpha (α) and composite reliability (CR) measures are well above 0.7, indicating construct level reliability. Second, the Fornell and Larcker approach (Fornell & Larcker, 1981) indicates convergent validity as the average variance extracted (AVE) for each construct is well

above .50. Moreover, as the average variance extracted (AVE) for each construct exceeds the squared correlations between all pairs of constructs, we can assume discriminant reliability (Hulland et al., 2018).

Table 16. Measurements

Construct	Item	Loading
Interface Trust (Srivastava & Chandra, 2018) $\alpha = .96$ AVE = .76 CR = .78	I believe this virtual assistant to be trustworthy.	.871
	I consider this virtual assistant to be reliable.	.865
	I trust this virtual assistant.	.877
Service Satisfaction (Bettencourt, 1997) $\alpha = .96$ AVE = .72 CR = .84	Overall, I am satisfied with TER Insurance.	.852
	The performance of TER Insurance has fulfilled my expectations.	.834
	TER Insurance compares to my vision of an ideal insurance company.	.855
Firm Benevolence (Schlosser et al., 2006) $\alpha = .94$ AVE = .61 CR = .78	I believe that TER Insurance would act in my best interest.	.776
	If I required help, TER Insurance would do its best to help me.	.793
	TER Insurance is interested in my well-being, not just its own.	.780

C.4 Results

First, to test H_1 and H_2 a two-way ANCOVA was conducted to assess the effect of the voice gender condition and the conversational style condition on interface trust controlling for participants' demographics. The results did not show a significant main effect of the voice gender condition on interface trust ($M_{\text{Male}} = 4.92$, $M_{\text{Female}} = 5.08$, $F = .717$, $p = .39$). Moreover, the main effect of the conversational style condition on trust was not significant ($M_{\text{TaskOriented}} = 5.08$, $M_{\text{SocialOriented}} = 4.90$, $F = .984$, $p = .32$). However, we found a significant interaction between both experimental factors ($F = 5.253$, $p < .05$). Follow-

up contrasts revealed that, there was no effect of voice gender on interface trust in the task-oriented conversational style condition ($M_{\text{TaskOriented/MaleVoice}} = 5.19$, $M_{\text{TaskOriented/FemaleVoice}} = 4.96$, $t = .856$, $p = .39$). Conversely, we found that participants' experience of interface trust was increased for the female voice condition for the social-oriented conversational style ($M_{\text{SocialOriented/MaleVoice}} = 4.61$, $M_{\text{SocialOriented/FemaleVoice}} = 5.21$, $t = -2.037$, $p < .05$). Additionally, we found that interface trust was decreased for the social-oriented conversational style in the male voice condition ($M_{\text{MaleVoice/TaskOriented}} = 5.18$, $M_{\text{MaleVoice/SocialOriented}} = 4.60$, $t = 2.136$, $p < .05$). There was no significant effect of conversational style on voice gender in the female voice condition ($M_{\text{FemaleVoice/TaskOriented}} = 4.96$, $M_{\text{FemaleVoice/SocialOriented}} = 5.21$, $t = -.857$, $p = .39$). Taken together, these results refute H_1 , but provide support for H_2 . Figure 3 illustrates the interaction effect of voice gender and communication style on interface trust.

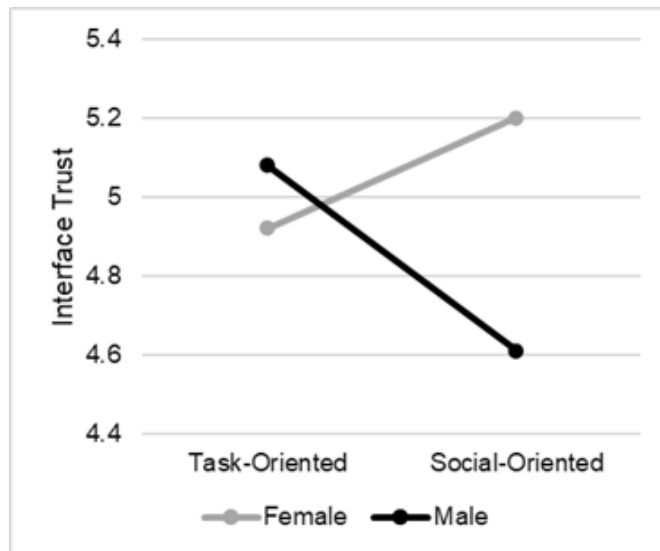


Figure 14. Interaction of Voice Gender and Conversational Style

To assess the overall system of hypotheses, we calculated two moderated mediation models (A. F. Hayes, 2017), 5,000 bootstrap samples, model 7) with each service satisfaction and firm benevolence as the focal dependent variable. The voice gender condition served as the independent variable, interface trust as the mediating variable, and the conversational style condition as a moderator of the effect of the voice gender condition on interface trust. As shown in Figure 3., these findings demonstrate that there is no direct effect of voice gender on interface trust ($\beta_{\text{Male}} = .22$, $SE = .27$, $t = .835$, $p = .40$). However, we find that the conversational style manipulation significantly moderated the effect of voice gender on interface trust ($\beta_{\text{Male} \times \text{SocialOriented}} = -.83$, $SE = .39$, $t = -2.107$, $p < .05$), indicating that the male voice paired with a social-oriented conversational style

decreased interface trust. Furthermore, when using service satisfaction as the focal dependent variable, we find that interface trust significantly affects service satisfaction ($\beta_{\text{Trust}} = .79$, $SE = .05$, $t = 15.548$, $p < .001$). The model led to a significant moderated mediation index excluding zero ($\beta_{\text{Indirect}} = -.65$, 95% CI: [-1.29, -.03]). Moreover, when using firm benevolences as the focal dependent variable, we find that interface trust also significantly effects firm benevolence ($\beta_{\text{Trust}} = .69$, $SE = .06$, $t = 11.26$, $p < .001$). Again, this model led to a significant moderated mediation index excluding zero ($\beta_{\text{Indirect}} = -.56$, 95% CI: [-1.13, -.03]). These results show that greater levels of interface trust led to a more positively-valenced service experience and attribution of benevolence to the firm, thus providing support for H_3 . Moreover, these results indicate that interface trust mediates the effect of the interaction of voice gender and conversational style on service satisfaction and firm benevolence, such that pairing a male voice with a social-oriented conversational style negatively impacts interface trust and firm perceptions.

C.5 Discussion and Contribution

VUIs continue to grow in importance for service provision across the customer journey. Many of these interfaces utilize a female voice. How customers perceive voice gender can influence the trust in the interface and perceptions of the respective firm. Hence, understanding user perception of gendered VUIs is integral to their adoption and customer satisfaction. Previous research has shown that the impact of voice gender is highly contextual. However, no prior research has systematically investigated the combined effect of voice gender and conversational style on interface trust and firm perceptions. In an online experiment with 211 participants, we investigate the nuanced effects of voice gender and conversational style on user experience when filing an insurance claim. The findings of this research demonstrate that voice gender affects interface trust and thus firm perceptions only in combination with the conversational style used by the interface. Specifically, we show that a male voice is trusted less when paired with a social-oriented conversational style, ultimately leading to a more negatively-valenced service experience and less attributions of benevolence toward the firm. Interestingly, we do not find any significant effects of voice gender on participants' trust in the interface for a task-oriented conversational style. This indicates that in the context of filing an insurance claim, gender is not the sole factor influencing interface trust but rather that the gender conforms to role expectations manifested in the conversational style. Moreover, one possible explanation could be that the task-oriented conversational style reduces the tendency to anthropomorphize the interface, ultimately decreasing the likelihood to apply gender-stereotypes to VUIs.

Our results make two novel contributions to research on trust in automated service interfaces and the technology-automated customer. First, to the best of our knowledge, we are the first to assess the combined effect of voice gender (female vs. male) and conversational style (social-oriented vs. task-oriented) on interface trust in a fully conversational voice-mediated service interaction. In this context, we extend and build on previous work (e.g., Nass et al., 1997; Tay et al., 2014), showing that the effect of voice gender on interface trust is highly context-dependent. This work introduces and empirically tests a novel boundary condition that can impact how voice gender affects interface trust. Specifically, we demonstrate that using a less formal conversational style (i.e., social-oriented) affects interface trust such that male voices are perceived as less trustworthy, ultimately leading to less positive firm perceptions. However, we find no effect of voice gender on interface trust for a task-oriented conversational style. These findings provide a new look at component fit theory (Isbister & Nass, 2000), showing that the interplay of different design elements of a VUI can provide new insights beyond the individual consideration of these elements. Specifically, this work underlines the importance of creating consistent VUI characters. As such, we provide further evidence for the critical importance of the theory-driven “conversational design” of VUIs.

Second, we stipulate novel insights into user-technology relationships (Hoffman & Novak, 2018; Schuetz & Venkatesh, 2020) and how the interaction with anthropomorphized computer interface creates a new “assemblage” of user experiences. Specifically, this research provides evidence that the influence of the design of VUIs reaches beyond the assemblage itself (i.e., the trust relationship between the user and the interface). It also impacts the entity providing the interface in terms of evaluating the overall service experience and the firms’ benevolence.

These findings have immediate implications for firms employing VUIs to automate selected service encounters. This work suggests that the design features of a VUI must be carefully orchestrated to create trustworthy user experiences. Current industry practice in designing VUIs is often driven by trial and error. Few companies strategically map voice features (such as gender) onto the characteristics of the conversation itself (e.g., conversational style). Specifically, our results suggest that pairing a male voice with a social-oriented style harms user experience. Moreover, the result of this study indicates that a task-oriented conversational style is less prone to trigger gender-based stereotyping. At the same time, we do not find immediate benefits in using a social-oriented conversational style in terms of interface trust. As such, especially in sensitive contexts,

firms could be advised to use a task-oriented conversational style to create more inclusive user experiences and to avoid adverse effects based on stereotyping.

C.6 Limitations and Future Research

This study faces several limitations that accentuate the need for future research. First, this study is limited to the context of insurance claims, which represents a highly standardized and formal service encounter. However, as this study shows, the perception of voice gender is highly contextualized. Future research could investigate the interplay of voice gender and conversational style in different contexts (i.e., contexts that require great empathy, such as health-related tasks). Second, although vignette studies have proven reliable results (Aguinis & Bradley, 2014), our results are limited by only investigating perceptual service outcomes. Hence, future work should investigate whether voice gender (and its interplay with other design features such as conversational style) affect trust-related, consequential service outcomes such as interface usage or customer loyalty. Finally, we have only looked at the effect of a binary gender operationalization (e.g., female vs. male). However, technology firms such as Amazon and Google are working on implementing gender-ambiguous voices. This could provide more inclusive user experiences and prevent the reinforcement of harmful gender stereotypes (Sutton, 2020). Hence, future research could investigate how consumers evaluate gender-ambiguous voices in customer service settings and whether these voices effectively create a more inclusive user experience in an increasingly AI-mediated customer journey.

Paper D – Female by Default? – Exploring the Effect of Voice Assistant Gender and Pitch on Trait and Trust Attribution

Table 17. Bibliographical metadata Paper D

Title	Female by Default? – Exploring the Effect of Voice Assistant Gender and Pitch on Trait and Trust Attribution
Authors	Tolmeijer, Suzanne; Zierau, Naim ; Janson, Andreas; Wahdatehagh; Leimeister, Jan Marco; Bernstein, Abraham
Publication outlet	<i>CHI 21 Extende Abstracts</i>
Publication type	Conference Paper
Publication year	2021
Rating (Jourqual 3)	Not applicable
Publication status	Published

Abstract

Gendered voice based on pitch is a prevalent design element in many contemporary Voice Assistants (VAs) but has shown to strengthen harmful stereotypes. Interestingly, there is a dearth of research that systematically analyses user perceptions of different voice genders in VAs. This study investigates gender-stereotyping across two different tasks by analyzing the influence of pitch (low, high) and gender (women, men) on stereotypical trait ascription and trust formation in an exploratory online experiment with 234 participants. Additionally, we deploy a gender-ambiguous voice to compare against gendered voices. Our findings indicate that implicit stereotyping occurs for VAs. Moreover, we can show that there are no significant differences in trust formed towards a gender-ambiguous voice versus gendered voices, which highlights their potential for commercial usage

Keywords

Voice Assistants, Gender Stereotypes, Voice Design, Trust, Gender-Ambiguous Voice

D.1 Introduction

Voice Assistants (VA), such as Google Assistant or Amazon Alexa, promise to change the ways people perform tasks, use services, and interact with organizations. The interactions of many users with these agents, however, have yielded mixed results, indicating high failure rates (Fuckner et al., 2014). Hence, there has been a growing interest in voice-based interactions in both research and practice (Riccardi, 2014). Besides the content of the interaction itself (i.e., “what is said?”), an element that is central to interaction design of VAs is the voice (i.e., “how it is said?”) (Sutton, 2020). In this regard, a prevalent trend is the use of female over male voices, as companies cite anecdotal evidence which suggests that female voices are favored by most users. Thus, most leading VAs are exclusively female or female by default (Loideain & Adams, 2020). In fact, according to a recent study, 77\% of all virtual assistants manifested gender-specific cues that can be classified as feminine (Feine et al., 2019). However, a recent report by the UNESCO stresses that the gendered design of most VAs could solidify harmful gender stereotypes (UNESCO, 2019). For instance, since people become used to interacting with those agents in a commanding tone, humans might also (subconsciously) mirror this behavior in their everyday conversations with women (Chin et al., 2020). One potential solution to this issue may lie in the use of gender-ambiguous² voices (UNESCO, 2019). Gender-ambiguous voice assistants may not only help to combat hurtful gender stereotypes, but also provide more inclusive design tools to represent voices outside the binary gender identities. However, while studies on interaction design with VAs are growing (e.g., Myers et al., 2018; Parviainen & Søndergaard, 2020; Purington et al., 2017; Zierau, Elshan, et al., 2020), there is a lack of empirical insights on the perceptual effects of gendered (and gender-ambiguous) voices based on para-lingual cues such as pitch. Especially, to the best of our knowledge, no study has empirically tested user perceptions and the technical feasibility of deploying gender-ambiguous voices for VA design.

To address this shortcoming, we conducted an exploratory study to empirically analyze the effects of (ambiguously) gendered voices on trait and trust attribution across different task contexts. Specifically, we comparatively analyze user perceptions in regards to pitch (low, high) and gender (female, male) as well as a gender-ambiguous voice we

² In accordance with Sutton (2020), we use the term “gender-ambiguous” throughout this paper rather than calling a voice ‘genderless’: many cues in the sound and content of VA speech can illicit gender ascription, even when the pitch is gender-neutral.

constructed. According to literature, the pitch of the voice is one of the most important factors regarding the attribution of gender (O'Connor & Barclay, 2017). To that end, we developed a voice interface for online experiments. On this basis, we implemented two task scenarios: one where users were asked to book a flight with a VA (assistance scenario) and one where users were surveyed by a VA on their financial situation (compliance scenario). We conducted a 5x2 online experiment with 234 participants on Prolific: five voices (male-low, male-high, gender-ambiguous, female-low, and female-high) were set against two task settings (assistance and compliance). Our results show implicit stereotype activation with regards to (lack of) trait attribution towards the different VA voices. Task context and gender of the participant both have an effect on perceived traits and reported trust. Finally, our study gives a first indication that a gender-ambiguous voice for VAs could be a viable alternative to gendered voices and warrants further investigation.

D.2 Related Work

Our research is motivated by sociophonetics and social response theory (Nass et al., 1994). Every person has a unique voice based on a complex interplay of anatomical and psychological traits and emotional states that together determine how people express themselves verbally and in turn how they are perceived by others (Hildebrand et al., 2020; Sutton et al., 2019). Sociophonetics explores how different speech patterns vary across social categories and the associated socio-cultural assumptions they carry. It is well established that people make inferences on others based on the sound of their voice (O'Connor & Barclay, 2017). Voice carries para-lingual cues that allow people to make assumptions about a person's background and, based on this, to apply social stereotypes. Speakers use subtle para-lingual cues, mostly unconsciously, to induce certain images to listeners (Sutton et al., 2019). Those cues can be seen as a flexible resource that people (and VAs) can use to signal different social traits and attitudes (Sutton et al., 2019). Sometimes, voice informs stereotypes about how specific groups of people speak. One obvious group is the gender of the speaker. The most prominent gender-dependent feature of voice is the pitch of a voice. The longer and thicker vocal chords of men produce a lower pitch than woman; a distinction that is easily perceived by listeners (O'Connor & Barclay, 2017).

Based on the Computers As Social Actors (CASA) paradigm (Nass et al., 1994), initial research suggests that when applied to technology, gender-specific voice characteristics may evoke stereotypical trait inferences (E. J. Lee et al., 2000). While this is not always

consciously, it is shown to be the case on a subconscious level (Mitchell et al., 2011). For instance, Pak et al. (2014) showed that users apply gender stereotypes when ascribing the trustworthiness of a virtual agent (i.e., the authors found that users trust a male more than to a female virtual doctor). In VAs, we find similar results. Initial findings suggest that people find it easier to process stereotypical voices, i.e., a warm gentle female voice and an assertive, forceful male voice (Strand, 2000). Specifically, it was shown that the machine's synthetic voice pitch can activate gender stereotyping of users. For instance, Nass et al. (1997) demonstrated that participants not only attributed gender towards computers that communicated in a low- versus a high-pitched synthetic computer voice. They also showed that the low versus high pitch of the synthetic voice triggered users to apply gender-schematic judgments of the “male” versus the “female” computer (Nass et al., 1997). More recently, Yu et al. (2019) found in their study that participants were more likely to disclose personal information to a (lower-pitched) male voice than a (higher-pitched) female voice of a virtual assistant (Yu et al., 2019). However, research that systematically analyzes trait and trust attribution based on different voice genders and pitches for VAs is scarce, despite its paramount role in VA design.

Another limitation to our understanding of voice pitch perception is that only male and female VA voices have been explored, despite calls to research a gender-ambiguous voice (Sutton et al., 2019; UNESCO, 2019). There is very little literature available on a gender-neutral voice pitch, except for references to “Q”, a voice that was recently created to be used for VAs to circumvent stereotyping (Q, 2019). The creators of “Q” mention the fundamental frequency should be between 145 and 175Hz for the voice to sound gender neutral. However, they indicate that gender is more than just pitch: tone and harmonics (e.g., the sound of vowels) also influence gender perception (Dennison, 2006). As the term gender-ambiguous indicates, voice cannot be regarded as binary (Baird et al., 2018; Sutton, 2020). A brain activity study done by Junger et al. found that people have an increased brain response to gender-ambiguous voices and opposite gendered voices cause stronger activation in the fronto-temporal neural network (Junger et al., 2013). While the difference in neural perception is shown, the difference in user perception for VAs has not been investigated. The use of gender-ambiguous voices, if proven not to have a negative impact on user trust and experience, can be a viable alternative to gendered voices to create a more inclusive environment for non-binary voices.

D.3 Method

In order to investigate the perceptual effects of voice pitch, we conducted a 5x2 between subjects experiment that manipulated i) the VA's voice gender based on pitch and ii) the task context the VA was deployed in. Dependent variable measures included trait ascription and reported trust in the VA. Specifically, in this exploratory study, we investigated the following research questions: *RQ1*: How does voice gender based on pitch affect trait ascription? *RQ2*: How does voice gender based on pitch affect user trust? *RQ3*: How does the task context influence the way how voice gender based on pitch affect trust and trait ascription?

D.3.1 Experimental Platform and Voice Design

The experiment was executed in a custom developed online voice assistant interface. By keeping the interface constant and as clean as possible, the focus remains on the voice of the voice assistant (see Figure 15), which allows to investigate trait attribution based on voice characteristics. The interaction with the VA follows a simple turn-taking mechanism, where the VA guides the unfolding conversation with the user. After each utterance of the VA, the button 'record' appears to send the user's response to the server. To control for diversity in the conversation, VA responses were prerecorded and the conversation path was delimited to focus on the task at hand.

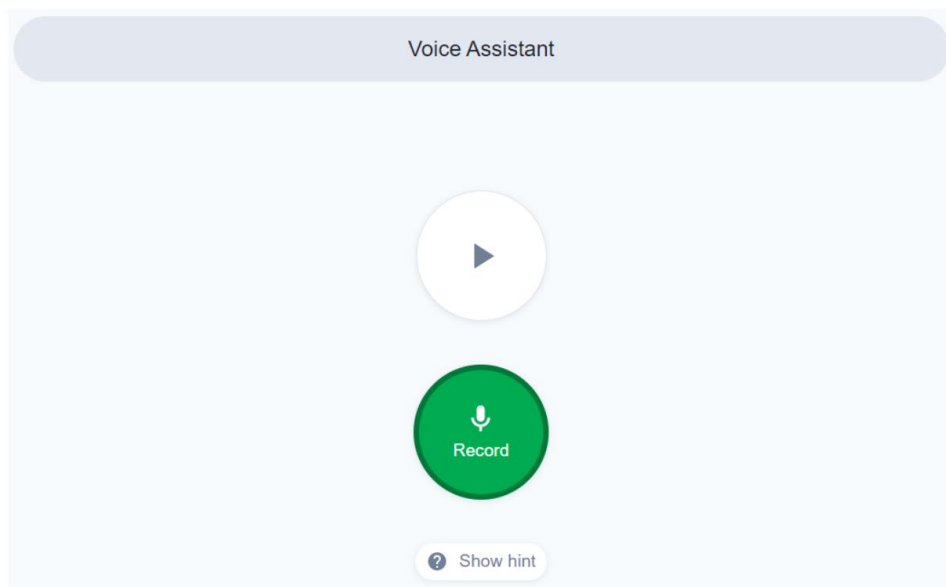


Figure 15. Online voice assistant interface

The prerecorded answers of the VA use state of the art in text-to-speech generation to produce our voice responses: Google WaveNet (Oord et al., 2016). To account for both gender and pitch differences, five American English voices are selected: a high- and

low-pitched female voice (based on voice *en-US-Wavenet-F*), a high- and low-pitched male voice (based on voice *en-US-Wavenet-B*), and a gender ambiguous voice (based on voice *en-US-Wavenet-E*). While gendered voice generators are readily available, there is not yet a gender-ambiguous text-to-speech generator available. Google's text to speech generator has it listed as an option that is not yet supported³. The only available gender-ambiguous generated voice is a carefully crafted voice clip called 'Q', created to fight gender stereotypes in voice assistants (Q, 2019). But "Q" offers no text-to-speech generation. In order to create a voice closest to gender-ambiguous, we pretest male voices with their pitch shifted up, and female voices with the pitch shifted down to identify a voice that classifies as gender-ambiguous. In this regard, gender-ambiguous refers to a voice that falls into both spectrums, meaning that different people would assign different genders to it based on prior mental models. Research on third gender associations has shown that typically people assign a gender to a voice, even though they cannot intuitively assign a gender (Sutton, 2020). To account for this tendency, we included a survey measure asking respondents to identify the gender of the voice assistant through three categories: (1) female; (2) male; and (3) unsure. We used this as a control measure in our models. Eleven manipulated voices based on different Google WaveNet voices were pretested by 52 participants on Prolific (47% female, average age 45 and ranging from 27 to 74). The voice receiving the highest division between assigned gender (58% male and 42% female) was voice *en-US-Wavenet-F* shifted down by three semitones. The selected voices can be found in Table 18.

³ When last checked by the authors on December 11th 2020. <https://cloud.google.com/text-to-speech/docs/reference/rest/v1/SsmlVoiceGender>

Table 18. Overview of selected voices

Voice Type	Female high (FH)	Female low (FL)	Gender-Ambiguous (A)	Male high (MH)	Male low (ML)
Original English US Wavenet voice	F	F	E	B	B
Method	TTS	TTS	GEN	TTS	TTS
Semitones pitch shift	+2	-6	-3	+2	-6
Average pitch	235 Hz	150 Hz	141 Hz	162 Hz	106 Hz
Original Google English US Wavenet voices are shifted by amount of semitone, either using Google's text to speech API (TTS) or by using online generator https://onlinetonegenerator.com/ (gen).					

D.3.2 Experiment Procedure

The experiment consisted of three phases: 1) randomization, 2) experimental task, and 3) post-test. Randomization and post-test were constant for all groups. Two different experimental task types were used: an assistant and a compliance task. These tasks are inspired by classical gender stereotypes: women are considered to be better in an assistant role, while men are more likely to be seen as leaders (Gustavsson, 2005; Koenig et al., 2011). Additionally, they are realistic VA tasks, as both customer surveys (Vernuccio et al., 2020) and assistant tasks (Loideain & Adams, 2020) are currently used in VAs. The assistance task involves booking a flight. The participant is given details for a specific flight they want to book and the VA will ask them questions to find and book the right flight for them. The compliance task focuses on personal questions asked in the context of a customer survey. People are asked to answer the questions, but are told it is possible to skip the answer if they prefer not to answer. An example of task interactions can be found in Figure 16.



Figure 16. Example excerpt of the compliance task

D.3.3 Participants

Participants were approached on crowdsourcing platform Prolific⁴. While complying with academic and Prolific's standards on data collection, we set the following preconditions: 1) US nationality, 2) 75%+ approval rate, 3) 10+ previous submissions, and 4) not in pretest sample. Requirement 1) was implemented to control for a language/culture barrier, as the selected voices are speaking in US English. Requirement 2) and 3) were applied to have some quality control in our sample. Requirement 4) excluding priming or bias stemming from the pretest. Initially, 345 people participated. We excluded participants who did not complete the entire task or failed the attention test. After data cleaning, we were left with 234 participants (96 male). The average age was 33 years old, ranging from 19 to 74.

D.3.4 Measurement and Analysis

The assignment of traits was measured by asking participants about the presence of 24 traits of the VA, based on male and female stereotypes (Bem, 1981; Eyssel & Hegel, 2012). Each trait was enquired using a 5-point Likert scale, ranging from a positive trait ascription (i.e., 5 indicates “*strongly agree*” that the VA had this trait), to negative trait ascription (i.e., 1 indicates “*strongly disagree*” that the VA had this trait). Female traits were averaged to indicate female stereotype activation ($\alpha = 0.91$), the mean of male traits was used to indicate male stereotypes ($\alpha = 0.87$). Perceived trust was measured using a validated questionnaire about the perceived competence, benevolence, and integrity of the VA ($\alpha = 0.93$) (Wang & Benbasat, 2005).

⁴ <https://www.prolific.co/>

D.4 Results

Trait ascription scores were not normally distributed: a Shapiro-Wilkin test resulted in $p < 0.001$ for all twenty-four traits. This is possibly because of the nature of the Likert scale for trait ascription: “1” indicates “strongly disagree” that the VA has this trait, “3” shows the participants “neither agrees nor disagrees”, while “5” reflects a “strong agreement” that the VA has this trait. To test whether a trait was significantly assigned in a positive way (i.e., significantly higher than a neutral answer of “3”) or a negative way (i.e., significantly lower than a neutral answer of “3”), we used the non-parametric paired sample Wilcoxon Signed-Rank test to compare our sample against the neutral value “3”.

Trust scores were not normally distributed either when comparing different voices: a Shapiro-Wilkin test showed the female high voice data was not normally distributed ($W = 0.95$, $p = 0.04$). As such, a Kruskal-Wallis test was used rather than an ANOVA test. In the case of two-group comparisons, Mann-Whitney U tests were executed. The results of all tests can be found in the remainder of this section.

D.4.1 Trait Ascription

While we found no significant activation of combined average male and female traits, results did show significant negative trait ascription. Specifically, over both task types, participants indicated that on average, some VAs did not have male and female traits. When taking the average over all stereotypically male traits, only the male low voice was not negatively marked as stereotypical male ($Z = 341$, $p = 0.175$). All other voices we significantly negatively associated with a male stereotype (MH: $Z = 268$, $p = 0.006$; FL: $Z = 461$, $p = 0.029$; FH: $Z = 198$, $p < 0.001$; A: $Z = 186$, $p = 0.006$). Female traits were only negatively assigned to low voices: the male low voice ($Z = 198$, $p = 0.006$) and female low voice ($Z = 532$, $p = 0.034$) were not considered to have stereotypical female traits. Other voices did not have negative stereotype ascription (MH: $Z = 456$, $p = 0.176$, A: $Z = 378$, $p = 0.543$, FH: $Z = 743$, $p = 0.903$). The gender of the participant did not influence negative stereotype assignment. The only voice that came near to activating a perceived stereotype was the female high voice: it was almost significant for activating a female stereotype ($Z = 743$, $p = 0.096$).

Additionally, we tested for group differences with regards to the individual perceived traits of the VA voices. Again, we added gender as a co-variate, to control for gender-specific differences in individual trait attribution. All voices were experienced to be organised, confident, cooperative, and polite. While low voices were overall considered

to be determined (ML: $Z = 257, p = 0.039$; FL: $Z = 759, p = 0.034$), only the low male voice was not experienced as friendly ($Z = 232, p = 0.250$). Curiously, a participant gender difference occurred in trait ascription to the gender-ambiguous voice: while all participants thought the voice was friendly and polite, women rated the ambiguous voice as significantly more friendly than men (*Mdn. women*: 5, *Mdn. men*: 4, $U = 84.5, p = 0.037$) and polite (*Mdn. women*: 5, *Mdn. men*: 4, $U = 65.5, p = 0.006$) than male participants. Two traits had no significant assignment of any kind for any voice pitch: assertive and affable. Interestingly, for many traits, the trait assignment was negative: people responded significant in the (strongly) disagree category. All voices were not considered to be aggressive, hard-hearted, tough, affectionate, sentimental, or romantic. However, implicit stereotype activation can be found in lack of negative trait ascription. For example, the low male voice was the only voice that was not considered *not* to be authoritative ($Z = 230, p = 0.356$) or dominant ($Z = 136, p = 0.057$). The female high voice, together with the gender-ambiguous voice, were the only voices not negatively assigned typical female traits such as delicate, family oriented, or sensitive. Difference in participant gender is more clear in negative trait ascription, as women assign lower values than men in many cases.

A summary of trait ascription can be found in Table 19, which shows trait ascription scores for all traits that were not uniformly assigned across all voices.

Table 19. Selected average trait ascription scores per voice pitch

	ML	MH	A	FL	FH
Authoritative	2.95	2.60*	2.18***	2.77**	2.51**
Speaks their mind	2.76	2.49**	2.74	2.46**	2.43**
Determined	3.21*	3.23	3.31	3.37*	3.25
Cold	2.98	2.77	2.49*	2.72*	2.27***
Dominant	2.73	2.51**	2.13***	2.32***	2.16***
Leadership skills	2.51**	2.74	2.69	2.53**	2.69*
Empathetic	2.19***	2.74	2.72	2.44***	2.90
Delicate	2.29***	2.47***	2.87	2.61*	3.12
Friendly	3.15	3.64***	3.59*	3.68***	3.98***
Sincere	2.98	3.13	3.49*	3.14	3.47
Family-oriented	2.54*	2.66*	2.87	2.63**	2.76
Sensitive	2.49**	2.57**	2.79	2.46***	2.84

*Selected average trait ascription scores per voice pitch. “1” implies strongly disagree the VA has this trait, “3” indicates the trait is not assigned, “5” shows a strong agreement for VA having this trait. Significant differences from lack of trait ascription, test by Wilcoxon Signed-Rank test, are shows as follows: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The five individual voices are male-low (ML), male-high (MH), female-low (FL), female-high (FH), and gender-ambiguous (A) respectively.*

D.4.2 Trust

Our results reveal no significant differences between the conditions when comparing reported trust in the VAs ($\chi^2(4, 234) = 1.9958, p = 0.736$). Average trust scores were comparable at 4.525 (ML), 4.480 (MH), 4.710 (FL), 4.636 (FH), and 4.824 (A). However, this does show that the gender ambiguous voice is not trusted less than gendered voices. Moreover, there is a significant difference in trust scores reported by male and female participants: female participants trust the gender-ambiguous voice more than men (*Mdn. women: 5.45, Mdn. men: 4.595, $U = 84.5, p = 0.048$*).

D.4.3 The Role of Task Context

In order to answer RQ3, we added task context as a variable in our analysis. For average male and female traits, context dependence is only seen for average male traits: the male low (*Mdn. assistant task (AT): 2.665, Mdn. compliance task (CT): 3.08, $U = 119, p = 0.028$*), male high (*Mdn. AT: 2.42, Mdn. CT: 3.0, $U = 177, p = 0.020$*), and gender-ambiguous voice (*Mdn. AT: 2.08, Mdn. CT: 2.83, $U = 113, p = 0.021$*) score higher on average male traits in the compliance task than the assistance task. Additionally, reported trust was stable over both tasks for male voices, while they were task dependent for female low (*Mdn. AT: 5.225, Mdn. CT: 4.18, $U = 250, p = 0.007$*), female high (*Mdn. AT: 5.18, Mdn. CT: 4.045, $U = 177, p = 0.003$*) and gender-ambiguous voices (*Mdn. AT: 5.045, Mdn. CT: 4.18, $U = 122, p = 0.038$*). In fact, all voices scored higher on trust for the assistance task compared to the compliance task.

D.5 Discussion

This study found evidence for the influence of voice gender and pitch on (stereotypical) trait attribution. While no positive stereotype activation was found, negative stereotypical trait ascription, and the lack thereof, showed implicit activation of gender stereotyping. For example, while the low male voice was not explicitly considered to be stereotypically male, only the low male voice was not perceived *not* to be typically male, and only the low voices – both male and female – were not refuted to be stereotypically female. As for trust attribution, we did not identify direct effects of voice pitch. However, a trend showed higher trust in the gender-ambiguous voice for female participants. Finally, task context influences both stereotype activation (for male traits) and trust (for female and gender-ambiguous voices).

With regards to trait attribution, our findings show mixed results with respect to the CASA paradigm. Negative trait ascription was prevalent, which can be both due to a lack of a perceived trait or a lack of viewing the voice as a social actor all together. While active stereotype activation was missing, the absence of stereotype negation seems to indicate an implicit gender bias. The fact that male traits and trust in female (and gender-ambiguous) voices was context dependent, indicates voice pitch and voice gender does subtly influence perception. With regards to trust formation, our results do not seem in line with prior research on the effect of pitch in inter-personal interactions, which indicate that people generally trust people with high-pitched voices more (O'Connor & Barclay, 2017). This may indicate that some of those mechanisms may be weaker in human-computer interaction. However, it has to be noted, that the means, especially

for voices with the particularly low and high voice, reveal a trend towards higher-pitched voices being trusted more. As our sample size is comparatively small, those results may become significant with a more appropriate sample size.

Task context did have an effect on perception and stereotype activation. Male voices were perceived more stereotypically male in a more “male” context of a compliance task. Female voices on the other hand were significantly more trusted in assistance tasks when compared to a compliance task. Curiously, these effects were both present for the gender-ambiguous voice: perceived male traits and trust were context dependent. While it is a positive indication that the gender-ambiguous voice was not assigned one specific gender, it also shows a risk: because the voice does not fit one gender stereotype, it also does not fit one stereotypical response, making it sensitive to multiple possible responses.

Nevertheless, the gender-ambiguous voice showed no significant trust differences when compared to the gendered voices. This is a promising first result, as there is very little research on the impact of gender-ambiguous voices. The fear that lack of a mental model and added cognitive load due to unrecognizable sex of the voice negatively influences trust does not seem to be confirmed by our research. More research is needed into different contexts and different pitches to confirm that the gender-ambiguous voice does not have a negative impact on trust compared to gendered voices. Overall, the gender-ambiguous voice was found to be organized, confident, cooperative, and polite; just as the gendered voices. This seems to be an encouraging initial result for use of gender-ambiguous voices in VAs. The fact that women have a higher trust in the gender-ambiguous voice than men warrants further research.

Our study had some limitations that should be pointed out. First, for a quantitative study, our sample size is comparatively small due to the study's exploratory character. Second, the participants were asked to imagine the scenarios to be real-life, which may threaten the external validity of our results. Although our study included actual voice interaction, future work may reexamine the results in a field setting. Third, it should be noted that results only capture first impressions of the VAs. A longitudinal perspective on trait ascription and trust formation should be included in future work. Furthermore, three different Google WaveNet voices were used to create the voices used in our experiment. We did not control for other voice characteristics such as timbre and tone, which could have influenced our results.

Additionally, a limitation lies in the created gender-ambiguous voice. As indicated, voice gender does not only come from pitch, but also from language usage and intonation. We controlled for this as much as possible by testing different pitch shifts for different voices, but all voices were originally gendered. The gender-neutral voice clip called Q (Q, 2019) was recorded by using voices of people that neither ascribe to the male gender nor the female one. The lack of gender-ambiguous voice generators or text-to-speech tools hampers the research into the possibilities of such a voice.

D.6 Contribution and Future Work

Our study has several theoretical and practical contributions to prior work in HCI research on the use of VA in commercial settings, the role of para linguistic cues for trait attribution, and the effective design of a VA's "personality". To the best of our knowledge, this is the first line of systematic research demonstrating how variations in voice pitch induce gender-specific trait attribution towards the agent, how such attributions affect important perceptual downstream consequences such as trust, and how such changes are impacted by the task context. Moreover, we develop and comparatively evaluate a gender-ambiguous voice with promising first results.

Our findings show stereotype activation is not as clear-cut as one might expect, but appears as a lack of stereotype negation. This combined with the influence of the participants' gender and task context asks for a more in-depth examination into stereotype activation and perpetuation of VAs. Additionally, gender-ambiguous voices are a promising avenue of research for VA design, to strive for more inclusive design. However, there is currently a lack of tools providing gender-ambiguous voice generation. We call upon researchers and industry alike to focus on the creating of gender-ambiguous voice tools to be able to research and provide more inclusive and stereotype avoiding voices for VAs.

D.7 Acknowledgments

This work is supported by the Swiss National Science Foundation, Grant 192718. Andreas Janson acknowledges support from the basic research fund of the University of St.Gallen. We thank Prof. Leah Ruppanner for her valuable feedback.

Paper E – Voice Bots on the Frontline: Voice-Based Interfaces Enhance Flow-Like Consumer Experiences & Boost Service Outcomes

Table 20. Bibliographical metadata Paper E

Title	Voice Bots on the Frontline: Voice-Based Interfaces Enhance Flow-Like Consumer Experiences & Boost Service Outcomes
Authors	Zierau, Naim; Hildebrand, Christian; Bergner, Anouk; Busquet, Fram; Schmitt, Anuschka; Leimeister, Jan Marco
Publication outlet	Journal of the Academy of Marketing Science
Publication type	Journal Paper
Publication year	2022
Rating (Jourqual 3)	A, FT50
Publication status	Accepted for publication

Abstract

Voice-based interfaces provide new opportunities for firms to interact with consumers along the customer journey. The current work demonstrates across four studies that voice-based (as opposed to text-based) interfaces promote more flow-like user experiences, resulting in more positively-valenced service experiences, and ultimately more favorable behavioral firm outcomes (i.e., contract renewal, conversion rates, and consumer sentiment). Moreover, we also provide evidence for two important boundary conditions that reduce such flow-like user experiences in voice-based interfaces (i.e., semantic disfluency and the amount of conversational turns). The findings of this research highlight how fundamental theories of human communication can be harnessed to create more experiential service experiences with positive downstream consequences for consumers and firms. These findings have important practical implications for firms that aim at leveraging the potential of voice-based interfaces to improve consumers' service experiences and the theory-driven "conversational design" of voice-based interfaces.

Keywords

Voice-based Interface, Speech Interaction, Natural Language Processing, Artificial Intelligence, Media Richness, Flow, Service Research.

E.1 Introduction

The proliferation of voice-based interfaces driven by recent advances in artificial intelligence and natural language processing is radically transforming how consumers interact with firms along every touchpoint of the customer journey (Huang & Rust, 2018; Mende et al., 2019; van Doorn et al., 2017). Voice-based interfaces have emerged as a novel interaction paradigm, allowing firms to engage consumers in increasingly personal ways, almost as if they are talking to a human service provider (Hollebeek et al., 2021; Huang & Rust, 2021). From ordering a coffee with the Starbucks Barista bot (Perez, 2017) to speaking to Bank of America's virtual advisor Erica to conduct money wires (Fuscaldo, 2019). Voice-based interfaces increasingly support consumers in completing service requests more naturally compared to traditional text-based interface modalities (J. Singh et al., 2020; Wirtz et al., 2018). Indeed, recent industry reports indicate that the number of voice-enabled interfaces is rising exponentially (e.g., 8.4 billion voice-based interfaces by 2024, Smith, 2018b).

Such voice-based interfaces combine automatic speech recognition with natural language processing to enable a fully voice-mediated service experience (Diederich et al., 2022; Seaborn et al., 2021). The fact that consumers execute service requests through voice-based interfaces highlights that they are both increasingly more comfortable interacting through voice-based interfaces and that firms expanded the use of voice-mediated service channels across touch-points (e.g., 72% of US consumers have already interacted with voice-based interfaces in business settings according to a study by PwC; P. Hayes & Wagner, 2018). This burgeoning trend raises the question of how consumers respond to such voice-based interactions in customer service settings, how they perceive the firm providing such voice-based service exchanges, and what underlying psychological mechanism might explain these effects with which downstream effects on consumers and firms.

To answer these questions, the current work introduces a novel conceptualization of voice-based interfaces in marketing. Specifically, the current work examines how the affordances of voice-based interfaces affect customer experiences during self-service encounters, building on and integrating prior work on multiformat communication (Moffett et al., 2021) and media richness theory (Daft & Lengel, 1986). The findings of this research shed light on how affordances unique to voice-based interfaces (i.e., verbal

cues, channel synchronicity) enhance flow-like user experiences and how these experiences, in turn, shape consumers' perception of their service experience, the firm, and downstream behavioral outcomes.

In what follows, we first discuss related work on voice-based interfaces, delineate the key conceptual properties of voice-based compared to text-based interfaces, and develop a set of hypotheses on how voice-based interfaces alter perceptions of flow, firm evaluation, and behavioral service outcomes. We then present the results of four studies that were designed to test our theorizing. Finally, we discuss the theoretical and practical implications for the effective conversational design of voice-based interfaces and the implications for customer service in an increasingly AI-driven economy.

E.2 Theoretical Background

E.2.1 Related Prior Work on Voice-Based Interfaces

In what follows, we will first review related prior work on voice-based interfaces and highlight how they relate to the current research. As summarized in Table 21, four key factors help to understand the unique insight of prior work with respect to (1) the nature or focus of the voice modality (voice primarily as input vs. output of the task), (2) the type of examined boundary conditions, (3) whether outcomes focused more on perception as opposed to behavioral consequences, and (4) the underlying psychological mechanism of the effect.

Table 21. Review of relevant literature on voice-based interface modalities

Study	Task	Voice Modality Focus	Psychological Mechanism	Boundary Condition(s)	Outcomes		Key Findings
					Perception	Behavior	
Qiu and Benbasat (2005)	Receive product recommendation	Output (Website help interface)		Interface design (Avatar presence)	Interface trust		Voice output increases cognitive and emotional trust toward the interface. Avatar presence strengthens this effect.
Hess et al. (2009)	Receive product recommendation	Output (Website help interface)	Social Presence	Interface design (Agent extraversion)	Interface trust		Agent extraversion strengthens the positive effect of voice output on social presence.
Qiu and Benbasat (2009)	Receive product recommendation	Output (Website help interface)	Social Presence	Interface design (Avatar presence)	Interface trust, Usage intention		Voice output increases perceptions of social presence towards the interface. Avatar presence strengthens this effect.
Klesse et al. (2015)	Order product	Input (Vending machine)				Product choice	Spoken preference expression prompts more indulgent product choices.
Pagani et al. (2019)	Search product information	Input (Search engine)		User characteristics (Privacy concerns)	Platform engagement, Brand trust		Voice input reduces platform engagement and ultimately brand trust. Higher privacy concerns strengthen this effect.
King et al. (2021)	Search product information	Input (Search engine)	Action Orientation		Intention to purchase		Voice input decreases purchase intention. This effect is mediated by reduced action orientation.
Berger et al. (2021)	Provide customer review	Input (Non-conversational interface)	Deliberation		Firm interest	Emotionality of expression	Voice input leads to more emotional attitudes, which is driven by reduced deliberation.
Son and Oh (2018)	Access video-on-demand content	Input & Output (Pre vs. post adoption of smart speakers)		User characteristics (Emotional attachment)	Purchase, Content consumption		The adoption of smart speakers increases content consumption and leads to more hedonic product choices. The effects on product choice are strengthened by emotional attachment towards the interface.
Rzepka et al. (2021)	Receive service recommendation	Input & Output (Conversational agent accessed on smart speaker vs. mobile device)	Efficiency, Enjoyment, Cognitive Effort	Task characteristics (Goal directedness)	Service satisfaction		Smart speakers enhance efficiency and enjoyment, reduce cognitive effort and, thus, increase satisfaction. These effects are strengthened by the goal directedness of the task.
This research	File insurance claim	Input & Output (Voice-based vs. text-based claim filing interface)	Interface Flow	Conversational design (Semantic fluency, Conversational turns)	Service experience, Firm evaluation	Contract renewal, Consumer sentiment, Conversion rates	Voice-based interfaces lead to enhanced interface flow and in turn improved perceptual and behavioral service outcomes. This effect is moderated by the conversational design, such that higher semantic fluency and reduced conversational turns strengthen the positive effect of voice-based interfaces on interface flow.

First, the majority of prior work either examined voice-based interface modalities by conceptualizing *voice as an input* as the focal independent variable or by examining *voice as an output* of the task, i.e. either assessing the role of the user's own speech during a task (*voice as input*) or how an interface responds back (*voice as output*). For example, Klesse et al., 2015) examined consumers' voice as an input in how selecting a product through spoken as opposed to written responses leads consumers to select more indulgent product options (see also Son & Oh, 2018) showing similar effects on hedonic product choices using a voice-based interface). In contrast, prior work focusing on voice as an output examined predominantly questions related to how much people trust a recommendation that is delivered verbally (vs. in written form) by a voice-based interface (Qiu & Benbesat, 2005). Only few studies examined a fully conversational interaction that required both the voice input by the user and a verbal response by the system (Rzepka et al., 2021). Second, the type of boundary conditions typically examined either appearance-related factors (e.g., predominantly the presence of avatars; see Hess et al., 2009; Qiu & Benbasat, 2009) or broad consumer characteristics (e.g., consumers' general privacy concerns or emotional attachment tendencies; see Pagani et al., 2019; Son & Oh, 2018). Except for emerging work examining the task characteristics as possible boundary conditions (Rzepka et al., 2021), no prior work we are aware of examined the unique *conversational design* characteristics during an interaction such as the specific semantic properties of the language or how the conversation itself is structured. Third, the type of outcomes studied in earlier work focused predominantly on perception outcomes (such as trust or satisfaction measures; Hess et al., 2009; Pagani et al., 2019; Qiu & Benbasat, 2009; Qiu & Benbesat, 2005) as opposed to the evaluation of the firm or downstream behavioral outcomes (see Son & Oh, 2018) for an important exception). Finally, the type of mechanisms examined in earlier work illuminated important processes such as the extent of perceived social presence (Hess et al., 2009; Qiu & Benbasat, 2009) as well as fundamental cognitive processes (such as the extent of deliberation or how efficient the process was perceived overall; J. A. Berger et al., 2021; Rzepka et al., 2021). However, whether affordances unique to a voice-based interaction also change how consumers experience a service interaction, such as whether the interaction is perceived as more immersive or absorptive, and how these effects in turn shape downstream firm-related behavior is unclear.

As summarized in the review of Table 21, the current work examines a fully conversational service interaction experience (voice both as input and output), examines key conversational design characteristics (how the conversation is structured and the adaptation

of the language during the interaction), examining both perceptual and behavioral outcomes relevant for consumers as well as firms, and a novel mechanism illuminating the extent of absorption during the task. In what follows, we will first review the key properties of voice vs. text-based communication followed by our conceptual model and hypotheses.

E.2.2 Key Conceptual Properties of Voice-Mediated Communication

Our literature review revealed that voice-based interfaces have been associated with a more intuitive, positively valenced experience (Rzepka et al., 2021) that can impact preference construction (Klesse et al., 2015) and the type of content that consumers reveal about themselves (J. A. Berger et al., 2021). These effects raise the question about the underlying properties that make speaking versus writing truly unique. We build on Moffett et al.'s, (2021) framework of multiformat communication to carve out the unique nature of voice-mediated communication and how to separate the “cue characteristics” of verbal speech (more informal structure, flexible syntax, prosodic and temporal communication cues) and the channel characteristics (greater synchronicity due to fast-paced and more parallel processing with low revisability during a conversation, see Table 22)

Table 22. Key conceptual properties of text-mediated vs. voice-mediated communication

Communication Format	Text-Mediated Communication	Voice-Mediated Communication
Cue Characteristics	Textual - Formal structure and grammar - Precise syntax - Limited symbol set - No prosodic or temporal cues	Verbal - Informal structure - Flexible syntax - Extensive symbol set - Prosodic and temporal cues
Channel Characteristics	Low Synchronicity - Slower-paced - Asynchronous (sequential processing) High Revisability - Possibility to assess, deliberate, and rehearse	High Synchronicity - Faster-paced - Synchronous (parallel processing) Low Revisability - Little to no possibility to rehearse

Cue Characteristics. Verbal cues refer to the specific vocal features of spoken language (e.g., tone, pitch, inflection, accent; Hildebrand et al., 2020; Moffett et al., 2021; Walther, 2005), while textual cues are linked to the written or typed language, such as spelling and punctuation (Moffett et al., 2021). Prior work on media richness theory

suggests that the verbal cues in voice-based communication allow individuals to communicate more intuitively and naturally compared to text-based exchanges, leading to a sensorially richer experience (Daft & Lengel, 1986). For example, pauses, changes in speed, and differences in intonation evoke a more immersive experience during a voice-based as opposed to a text-based interaction (Redeker, 1984). This makes voice a richer and more vivid medium, leading to enhanced levels of involvement in a task (Daft & Lengel, 1986; Dennis & Kinney, 1998; Moffett et al., 2021; Steuer, 1992).

Channel Characteristics. The literature on the interactivity of media channels suggests that the speed and immediacy of an interaction with a medium can lead to an increase in its perceived interactivity (Steuer, 1992). Voice-based as opposed to text-based exchanges unfold more synchronously. This enhanced channel synchronicity in voice-mediated interactions has been shown to lead to enhanced absorption in a task (Agarwal & Karahanna, 2000; Haeckel, 1998; Moffett et al., 2021). Contrary to written communication that allows consumers to re-read and re-type responses (so-called channel revisability, Moffett et al., 2021), spoken language must be encoded and decoded in a parallel process, thus making voice-based interactions substantially more immediate (Rubin et al., 2000). Voice-based interfaces allow users to have a more immediate influence on the content and the process of communication, promoting the exchange of social information at a faster rate and even impacting early relationship development (Moffett et al., 2021).

E.2.3 Voice-Based Interfaces and Flow

How do these conceptual properties of voice-based communication (e.g., faster-paced, more parallel, synchronous processing of information and a rich sensory experience) impact consumers' task experience? One construct to describe such positively-valenced, immersive user experiences is the concept of *flow* (Agarwal & Karahanna, 2000; Csikszentmihalyi, 1975; Hoffman & Novak, 1996). Flow represents a pleasant, positively-valenced experience (Agarwal & Karahanna, 2000; Ghani et al., 1991; Trevino & Webster, 1992; Webster et al., 1993) that is characterized by a high level of immersion, positive affect, and optimal level of challenge (i.e., balancing boredom and excessively challenging experiences). Greater levels of flow have been shown to lead to more playful, intuitive, and effortless interaction experiences during web navigation (Hoffman & Novak, 1996) and a sense of total engagement and immersion with a system (Agarwal & Karahanna, 2000). Flow-like user experiences can even arise from the properties of the medium itself (Brannon Barhorst et al., 2021; D. Kim & Ko, 2019). For

instance, Brannon Barhorst et al. (2021) demonstrate that the unique affordances of virtual reality applications facilitate a state of flow through enhanced levels of interactivity and vividness of the experience. Due to the sensory-rich, faster-paced, and extensive symbol set of voice-mediated communication, it is plausible that these distinct features of verbal communication similarly affect the experience of flow in voice as opposed to text-based service interactions.

Flow is characterized by “a seamless sequence of responses facilitated by machine interactivity” and occurs when consumers are immersed in a computer-mediated activity (Hoffman & Novak, 2009). Thus, we expect that the extent of interactivity and vividness afforded by the medium will impact the extent of interface flow that users experience (Novak et al., 2000). Contributing to the expected impact of enhanced interactivity of the communication medium, prior work provides initial evidence that two-way interactions and synchronicity might impact flow experiences (Guan et al., 2021). Given that voice-based (as opposed to text-based) communication facilitates more immediate and synchronous interactions (Moffett et al., 2021), we expect that the greater richness of verbal cues and greater extent of synchronicity in voice-based interactions enhances consumers’ experience of interface flow, i.e. the extent of flow that consumers experience in connection with the interface. Thus, due to the unique cue and channel characteristics of voice-mediated communication, we formally hypothesize that using a voice-based interface (vs. text-based interface) enhances consumers’ experience of interface flow.

H₁: *Interacting with a voice-based compared to a text-based interface leads to enhanced levels of interface flow.*

E.2.4 The Impact of Interface Flow on Service Experience & Firm Outcomes

What might be the impact of enhanced levels of interface flow on service and firm outcomes? Prior work suggests that flow is a positively-valenced state that is intrinsically enjoyable (Agarwal & Karahanna, 2000; Hoffman & Novak, 1996; Johnson et al., 2003) and can even carry over to subsequent tasks, making them appear more experiential and enjoyable (Harmeling et al., 2017; Mathwick & Rigdon, 2004). Due to the increased sensory involvement and high synchronicity of the voice modality, we hypothesize that greater levels of flow during interface use carry over to consumers’ overall experience with the service. This is consistent with Zomerdijk & Voss’ (2010) work on designing “experience-centric services” that highlights the importance of the interface modality to provide more experiential service deliveries.

Experiences that go beyond customers' expectations have been shown to impact also the firm providing the service (Harmeling et al., 2017). Such experiences can move a customer from seeing interaction as merely transactional to building a positive consumer-firm relationship (Zomerdijk & Voss, 2010). This is consistent with prior research showing that interactive and multisensory service experiences can enhance firm perceptions along the customer journey (Lemon & Verhoef, 2016). In short, we expect that the enhanced experience of interface flow when using a voice compared to a text-based interface will in turn lead to an overall more experiential and positively-valenced service and firm evaluation.

H₂: *Greater levels of interface flow when interacting with a voice-based as compared to a text-based interface have a positive effect on (1) a more experiential service delivery and (2) a more positively-valenced evaluation of the service firm.*

These changes in perception might also impact consumers more directly. As the cue and channel characteristics of voice-mediated interfaces may drive a more experiential service delivery (H₂), it is plausible that they might also in turn help form a closer connection between the consumer and the firm (Harmeling et al., 2017; Schouten et al., 2007). While the experience of interface flow is transitory, the fact that it increases the enjoyment of an activity for its own sake suggests that it may have a more lasting impact on behavior related to future interactions with the same firm, making individuals wanting to experience a state of flow again and again (Schouten et al., 2007). Thus, we hypothesize that the enhanced level of interface flow that consumers experience when interacting with a voice-based interface will positively impact behavioral outcomes in connection with the firm. Specifically, we hypothesize that greater levels of interface flow in turn enhance consumers' desire to stay committed to their relationship with the firm in the future, to engage in future interactions with the service firm, and to advocate the firm and its services to others.

H₃: *More positive perceptions of the firm and the experiential service delivery (H₂), in turn, enhance consumers' contract renewal decision, consumer sentiment, and conversion rates to engage in further service exchanges with the same firm again.*

E.2.5 Boundary Conditions: The Role of Semantic Fluency & Conversational Turns

Flow has been found to only emerge when a user's skill matches the difficulty of the task (Hoffman & Novak, 1996). More specifically, there needs to be an optimal match between the capabilities of the user and the opportunities provided by the interface to

execute a task. While the literature vastly disagrees on the number of possible categories of combinations between different levels of user skill and task complexity (Clarke & Haworth, 1994; Csikszentmihalyi, 1988; Csikszentmihalyi & LeFevre, 1989) the literature largely agrees that for flow to emerge, there needs to be *some* level of challenge above a certain threshold (see prior work on optimal stimulation theory; Holbrook & Gardner, 1993; Steenkamp & Baumgartner, 1992). Thus, if the task is too challenging, users get anxious or frustrated, while if the task is not challenging enough, they may get bored. Thus, how a certain task is designed has an important impact on the extent to which flow can emerge.

This suggests that a key boundary condition to experience flow in CIs is linked to how the conversation is governed or designed. During interactions with a CI, the user engages in information processing based on *what* and *how* the interface communicates (H. H. Clark & Brennan, 1991; Pogacar et al., 2018). Thus, the way information is processed depends on both (1) how the language used by the interface is processed (what we refer to as the *semantic fluency* of a conversation), and (2) how the messages are structured as a whole (what we refer to as the extent of *conversational turns* during a conversation).

Semantic Fluency. Research in cognitive psychology and linguistics has demonstrated the important connection between linguistic properties of a conversation and the ease with which it is processed (Fernández-Sabiote & López-López, 2020; H. Song & Schwarz, 2008). Specifically, research has shown that simpler language leads to more fluent processing of messages (Thompson & Ince, 2013). We therefore expect that if an interface employs simpler language, users will process it with greater fluency. The differences in channel synchronicity between text-based and voice-based communication suggests that semantic fluency may differentially affect each channel. Indeed, research on interpersonal communication suggests that a simpler conversational style better suits the properties of a dynamic dialogue as opposed to those of written text (Redeker, 1984). This is primarily due to the extent of synchronicity of the communication channel. When using a text-based interface, consumers can re-read a message with lower semantic fluency to aid message processing due to the asynchronous nature of the channel. Conversely, when interacting with a voice-based interface, the synchronous nature of the channel makes processing low semantic fluency messages more difficult, requiring greater cognitive processing. Based on our theorizing about the informal structure, flexible syntax, and more extensive symbol set with low revisability in verbal compared to text-based communication (see Table 2 for a summary) as well as previous research showing that messages in spoken as opposed to written communication require greater

cognitive processing (e.g., van Zeeland & Schmitt, 2013), we expect that the cognitive processing of messages with lower semantic fluency is more pronounced in voice compared to text interfaces and will therefore affect flow more negatively in voice (as opposed to text-based) interfaces.

H_{4a}: The positive effects of voice-based interfaces on interface flow are increased (decreased) under conditions of high (low) semantic fluency.

Conversational Turns. The second boundary condition assesses to which extent consumers' experience of interface flow is disrupted by increasing the number of conversational turns during the interaction with a voice-based interface. Individual messages are linked together in a conversation by conversational partners taking turns, which can vary in length depending on the context of the conversation (Levinson, 2016). Thus, when interacting with a voice-based interface, users are forced to adjust to the pace (i.e. speaking rate) of the interface and remain attentive to what is being said throughout the conversation (Redeker, 1984). In contrast, when using a text-based interface, consumers can deliberately choose their own pace of processing information and rehearse or reflect as they wish on what was being said (Moffett et al., 2021). This suggests that the extent of conversational turns can either disrupt or enhance interface flow in voice-based interfaces. We expect that a greater number of conversational turns to complete a service task will reduce channel synchronicity and provoke a less absorptive experience, as the interface flow is interrupted more often. Indeed, as flow is experienced as a high level of continuous immersion in an activity (Csikszentmihalyi, 1975), we expect a higher number of conversational turns to cause a reduced absorption in a task. Thus, we expect that a higher number of turns creates a feeling of disjointedness between individual messages. Taken together, we predict that an increase in conversational turns will negatively impact interface flow in voice-based (vs. text-based) interfaces.

H_{4b}: The positive effects of voice-based interfaces on interface flow decrease (increase), with a greater (lower) number of conversational turns during the interaction.

Our overarching hypotheses are summarized in Figure 17. In what follows, we provide a short description of the experimental setup and technical implementation, followed by our empirical studies.

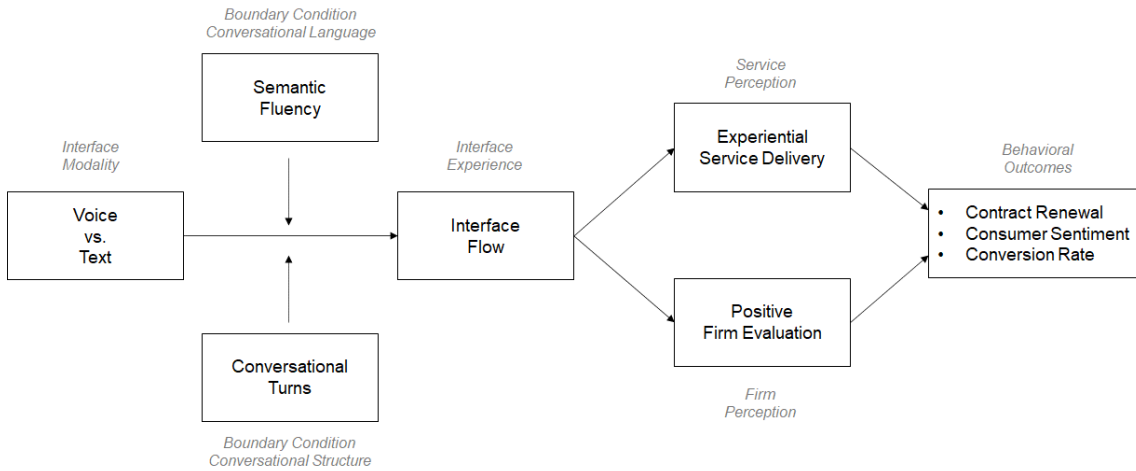


Figure 17. Conceptual model

E.3 Experimental Paradigm and Context

To test our hypotheses and to ensure maximum experimental control, we developed a custom-made voice-based interface (using Python and Google WaveNet; see Web Appendix A1 for details). The appearance of the interface was kept neutral to avoid any confounds that could have produced spurious effects on our key measures of interest (appearance of the avatar or any other visual design cues) and to isolate the hypothesized interface modality effects we predict. We provide a detailed description in the Web Appendix explaining the technical infrastructure for both research transparency reasons as well as to provide researchers with the opportunity to build on and expand on our current technical implementation. We predefined the dialogue flow based on the context of each study and ensured that the order of all questions was identical between conditions across experiments. We used Google’s state-of-the-art text-to-speech generator WaveNet for all studies (Oord et al., 2016) and participants went through a fully voice-based service interaction with individual prompts to complete the service task (see also Web Appendix A1 for further details). To rule out potential confounds related to the gender of the voice bot, we used both female (Study 1 and Study 4) and male voices across studies (Study 2 and Study 3). Finally, in determining our sample sizes, we followed prior work that used a similar experimental paradigm with intermediate effect sizes (e.g., King et al., 2021 sampled around 125 participants per condition comparing text-based vs. voice-based user search).

We focused on a service context with important economic and societal implications. Specifically, we developed a voice-based interface to submit and handle insurance claims for two major reasons. First, the claim filing process in the insurance industry has been heavily standardized to simplify claim management (from claiming the loss of

a mobile phone to reporting a car damage) (A. Singh et al., 2019). Second, the claim handling process is often perceived as a major source of distress for individuals that negatively affects their experience across service touchpoints (Riikkinen et al., 2018). According to recent industry reports, the trend to phase out human service agents in favor of using automated file claiming systems has also led to reduced customer satisfaction rates, enhanced client frustration, and an experienced loss of “human touch” (Blake, 2018), rendering the use of voice-based interfaces as a novel intervention in the current service automation landscape.

E.4 Overview of Studies

We present evidence from four studies that were designed to test our theorizing. In Study 1, we test our baseline hypothesis that a voice-based compared to a text-based interface promotes a more flow-like user experience, which in turn translates into a more positive service perception. Study 2 demonstrates an important boundary condition to the effect of voice-based interfaces, showing that reduced levels of semantic fluency can provoke negative user experiences which in turn also impact consumers’ inclination to extend their relationship with the service firm. Study 3 further demonstrates a second key boundary condition, showing that a greater number of conversational turns in a voice-based interface, reduces consumers’ experience of interface flow and perception of the service firm. Study 3 also rules out that the current findings can be explained by a general positive affect account. Finally, Study 4 further demonstrates the impact of voice-based interfaces on behavioral outcomes (conversation rates and consumer sentiment) and provides further evidence on the unique role of interface flow to explain these effects (consistent with optimal stimulation theory and the matching of task affordances and skills; Holbrook & Gardner, 1993).

E.5 Study 1

Study 1 was designed to provide a first test of our baseline hypothesis that a voice-based compared to a text-based interface alters perceptions of interface flow. Furthermore, the current study also tests whether these changes in interface flow promote more experiential service delivery. We also further explore whether these effects are conditional on consumers’ prior experience with voice technology.

E.5.1 Design and Procedure

A total of 184 participants ($M_{\text{Age}} = 32.97$, $SD_{\text{Age}} = 11.05$, 50% females; for an overview of the demographics across all studies see Web Appendix B1) were recruited from a

nationwide consumer panel (Prolific⁵) and randomly assigned to a two-cell between-subject experiment (text-based vs. voice-based interface). In both conditions, the experimental task comprised the filing of an insurance claim with a fictitious car insurance company called TER Insurance. Before the task, participants were presented with a scenario that described the events of an accident including the damage that occurred to the car (see Web Appendix A2.1 for a detailed description of the scenario). In the claim filing task, we asked participants to report all information concerning the incident via either the text- or voice-based interface (dependent on condition). The content and the sequence of questions were identical across conditions (i.e., the sequence of six conversational turns, content, and all other features of the file claiming process were held constant across conditions). Participants were screened to use a laptop or desktop computer (non-mobile devices) to avoid any interface modality confounds (such as differences in writing speed on mobile versus non-mobile devices). We used a female voice in this study but find consistent results across studies regardless of the gender of the voice-based interface (e.g., using a male voice in Study 2).

Measurement. Immediately after completing the insurance claim filing task, we assessed participants' perception of interface flow when using the service interface with three items (scale adapted from Unger & Kernan, 1983, sample item: "This interface totally absorbed me."; 7-point Likert scale, from 1: "Strongly disagree" to 7: "Strongly agree"; $\alpha_{\text{Flow}} = .93$, see Web Appendix A3.1). Next, we measured consumers' extent of experiential service delivery using a four-item scale (scale adapted from Unger & Kernan, 1983); sample item: "This interface offers novel experiences."; 7 point-Likert scale, from 1: "Strongly disagree" to 7: "Strongly agree"; $\alpha_{\text{Experiential}} = .94$, see Web Appendix A3.2). Finally, participants answered how frequently they use voice-based interfaces in general (see Web Appendix A3.5) and self-reported their domain-specific insurance expertise (see Web Appendix A3.4).

E.5.2 Results

Participants who used the voice-based interface experienced significantly higher levels of interface flow compared to participants using the text-based interface ($M_{\text{VoiceInterface}} = 3.09$, $M_{\text{TextInterface}} = 2.31$, $t = 2.996$, $p < .01$; Cohen's $d = .44$; see Figure 18). Similarly, we also found that participants using the voice-based interface reported a significantly

⁵ Peer et al. (2017, 2021) compared the data quality of different crowdsourcing platforms for behavioral research, showing that Prolific scored highest on response quality and sample diversity. Prolific was used to recruit participants and manage payments, and routed to our own server to complete the experimental task.

more experiential service delivery compared to participants using the text-based interface ($M_{\text{VoiceInterface}} = 3.78$; $M_{\text{TextInterface}} = 2.90$, $t = 3.239$, $p < .01$; Cohen's $d = .48$).

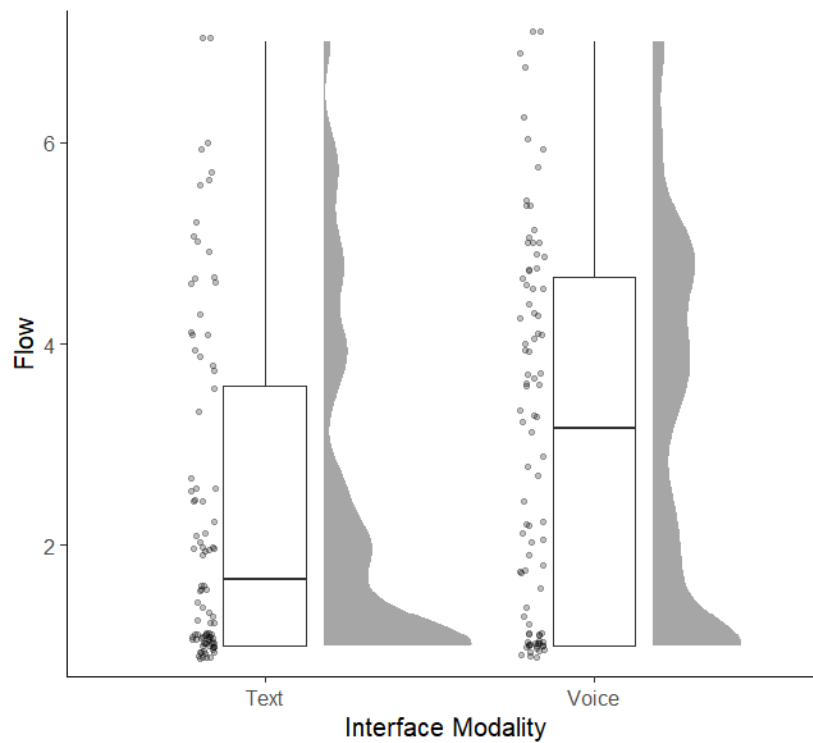


Figure 18. Voice-based interfaces promote flow-like consumer experiences

To provide a direct test of our theorizing on the effect of interface modality on consumers' perception of experiential service delivery, we estimated a simple mediation model (5,000 bootstrap samples) with the interface modality condition as the independent variable, interface flow as the mediator, and experiential service delivery as the dependent variable. In support of our theorizing, the voice-based interface led to a significant increase in interface flow ($\beta_{\text{Interface}} = .78$, $SE = .26$, $t = 3.002$, $p < .01$). In turn, greater levels of interface flow led to a more experiential service delivery ($\beta_{\text{Flow}} = .77$, $SE = .05$, $t = 14.649$, $p < .001$), rendering the residual direct effect non-significant ($\beta_{\text{Interface}} = .28$, $SE = .19$, $t = 1.469$, $p = .14$), and a significant indirect effect with a confidence interval excluding zero ($\beta_{\text{Indirect}} = .60$, 95% $CI: [.19, .99]$), indicating full mediation. These results were robust even after controlling for differences in participants' age, gender, income level, car ownership, domain expertise, and previous experience with voice technology (i.e., lower information criteria contrasting our main model to the alternative control model; $BIC_{\text{MainModel}}: 621.11$, $BIC_{\text{ControlModel}}: 637.80$). See Web Appendix B2 for additional statistical analyses including control variables.

E.5.3 Discussion

The findings of Study 1 provide initial evidence for the baseline hypothesis that voice-based interfaces evoke a more experiential service delivery compared to text-based interfaces and that these effects are driven by the flow-enhancing effect of voice-based interfaces.

E.6 Study 2

The objectives of Study 2 were threefold. First, the current study examines a theoretically important and practically relevant boundary condition. Study 2 tested to which extent the semantic fluency of a query can either enhance or reduce a users' perception of interface flow when using a voice-based interface. Second, we further test whether the use of a voice-based interface impacts only perceptual outcomes or also behavioral outcomes (contract renewal with the current service provider). Third, the current study used a male voice to rule out that the current findings are a function of the digital assistants' gender (female voice in Study 1 vs. male voice in Study 2).

E.6.1 Design and Procedure

A total of 612 participants ($M_{\text{Age}}=37.19$, $SD_{\text{Age}}=13.34$, 45.4% males) were recruited from a nation-wide consumer panel (Prolific) and randomly assigned to a 2 (interface modality: voice-based vs text-based interface) $\times$ 2 (semantic fluency: high vs. low) between-subjects design, using the same experimental paradigm and scenario as in Study 1. We used the same experimental paradigm as in Study 1 and only added short filler questions to provide a more realistic experimental setting (e.g., whether another insurance is already involved, and whether a technical expert was consulted; total of ten turns per interaction, see Web Appendix 2.2). We altered semantic fluency of each turn based on prior work in linguistics and communication research by replacing simple through more complex words (Khawaja et al., 2010; Redeker, 1984). For example, the high semantic fluency condition used easily to process phrases (e.g., "Please tell us more about the damage that you want to report."), the low semantic fluency condition used more difficult to process phrases (e.g., "Please elucidate on the nature of the damage that you wish to declare."). To avoid any confounding effects in terms of the length of a query and to isolate the effect of the semantic properties of the conversation, we held the number of words per turn constant between conditions ($M_{\text{HighFluency}} = 16.91$, $M_{\text{LowFluency}} = 17.09$, $t = -.04$, $p = .97$).

To study whether a voice-based interface might also impact behavioral outcomes, users were informed after completing the claim filing task that their contract is about to expire and that they could extend their contract if they wish to do so. All participants were then asked directly by the respective interface to either renew their contract or not (see Web Appendix A2.2 for further information on the scenario).

Manipulation Check. We used objective readability scores to assess the discriminatory power of the semantic fluency manipulation, using the Flesch reading ease score (FRE; higher is better / more fluent), the New Dale-Chall readability formula (NDR; lower is better / more fluent), and the Automated Readability Index (ARI; lower is better / more fluent). All measures confirmed that semantic fluency was indeed higher in the high versus low semantic fluency condition ($FRE_{HighFluency} = 80.03$ vs. $FRE_{LowFluency} = 56.44$; $NDR_{HighFluency} = 6.63$ vs. $NDR_{LowFluency} = 9.98$; $ARI_{HighFluency} = 3.15$ vs. $ARI_{LowFluency} = 6.69$)⁶.

Measurement. We measured consumers' experience of interface flow using the same items as in Study 1 ($\alpha_{Flow} = .92$), along with the same items to assess consumers' perception of experiential service delivery ($\alpha_{Experiential} = .91$). Moreover, we measured their perception of the service firm using a two-item scale (scale adapted from G. G. Lee & Lin, 2005; sample item: "Overall, I am satisfied with the online experience with TER insurance"; 7 point-Likert scale, from 1: "Strongly disagree" to 7: "Strongly agree"; $\alpha_{FirmEval} = .96$, see Web Appendix A3.3). Finally, we measured consumers' decision to extend their contract with TER insurance as 1 when extending and 0 when terminating their contract.

E.6.2 Results

Interface Flow. A two-way ANOVA revealed a significant main effect of condition on interface flow ($F(1, 608) = 46.84, p < .001, \eta^2 = .07$), demonstrating a significantly higher experience of interface flow in the voice as opposed to text-based interface ($M_{VoiceInterface} = 3.59, M_{TextInterface} = 2.72, t = 6.693, p < .001$). The semantic fluency main effect was non-significant ($F(1, 608) = 0.24, p = .63, \eta^2 = .00$) while the interaction between both experimental factors was statistically significant ($F(1, 608) = 6.72, p < .01, \eta^2 = .01$). Follow-up contrasts with Holm correction for family-wise errors revealed

⁶ To compute the FRE, the ARI, and the DCR we used the Spacy package in Python. In the web appendix (A3.6) we provide the corresponding formulas to compute these metrics. Please refer to DuBay (2004) for a review and methodological details.

that, consistent with our theorizing, participants' experience of interface flow in the voice-based interface condition was reduced in the low semantic fluency condition ($M_{\text{Voice_HighFluency}} = 3.77$, $M_{\text{Voice_LowFluency}} = 3.38$, $t = 2.189$, $p = .057$) while semantic fluency had no effect on participants in the text-based interface condition ($M_{\text{Text_HighFluency}} = 2.59$, $M_{\text{Text_LowFluency}} = 2.86$, $t = 1.470$, $p = .14$).

Service Experience. Mirroring the previous effect when assessing the impact on consumers' overall evaluation of the service, a two-way ANOVA revealed a significant increase of a more experiential service delivery in the voice as opposed to text-based interface ($F(1, 608) = 76.87$, $p < .001$, $\eta^2 = .11$; $M_{\text{VoiceInterface}} = 3.58$, $M_{\text{TextInterface}} = 2.51$), a non-significant main effect of semantic fluency ($F(1, 608) = 0.171$, $p < .78$, $\eta^2 = .00$), and significant interaction between both factors ($F(1, 608) = 6.365$, $p = .01$, $\eta^2 = .009$). Follow-up contrasts with Holm correction confirmed that participants in the voice-based interface perceived the service more negatively in the low semantic fluency condition ($M_{\text{Voice_HighFluency}} = 3.76$, $M_{\text{Voice_LowFluency}} = 3.40$, $t = 2.189$, $p = .07$) while semantic fluency had no effect in the text-based interface condition ($M_{\text{Text_HighFluency}} = 2.39$, $M_{\text{Text_LowFluency}} = 2.65$, $t = 1.470$, $p = .14$).

Path Model & Serial Mediation. Finally, we estimated a path model using the lavaan package in R with robust standard errors to assess the overall system of hypotheses and the impact of interface flow and service experience on firm evaluation and consumer choice (contract renewal). The path model results are summarized in Figure 19. These findings demonstrate that the increase of interface flow evoked by the voice-based interface ($\beta_{\text{Voice}} = .38$, $z = 6.504$, $p < .001$) in turn led to a significant increase in consumers' perception of experiential service delivery ($\beta_{\text{Flow}} = .84$, $z = 21.180$, $p < .001$) and an overall more positive perception of the firm ($\beta_{\text{Flow}} = .61$, $z = 15.320$, $p < .001$). Focusing on the impact on contract renewal, we found that the positive impact of the voice-based interface condition on service and firm perception, in turn, led to a significant increase in contract renewal rates ($\beta_{\text{ExpService}} = .16$, $z = 2.879$, $p < .01$; $\beta_{\text{FirmEval}} = .24$, $z = 4.615$, $p < .001$). These path model results were confirmed by a serial mediation model (Hayes 2017; model 81 with 5000 bootstrap samples) estimating the effect of the voice (vs. text) condition on contract renewal rate via flow (proximal mediator) and firm evaluation as well as service experience ratings (distal mediators), with a significant indirect effect excluding zero ($\beta_{\text{Indirect}} = .32$, 95% CI: [.16, .49]).

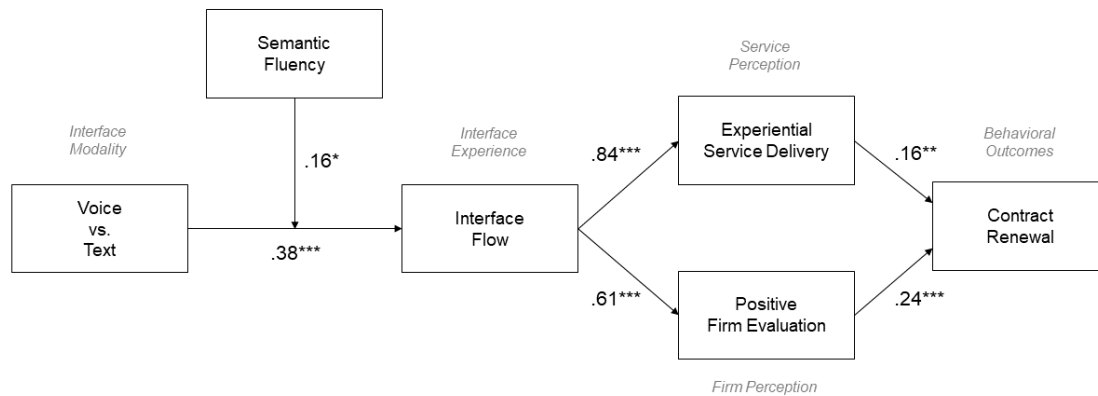


Figure 19. Path model results (Study 2)

E.6.3 Discussion

The results of Study 2 replicate the finding that voice-based interfaces impact consumers' experience of interface flow. The findings reveal further that these enhanced levels of interface flow drive both important perceptual outcomes (how consumers perceive the service firm and the service experience itself) as well as behavioral outcomes (consumers' decision to renew their contract with the same firm). We also provide evidence for an important boundary condition, demonstrating that these positive effects are reduced with lower levels of semantic fluency. While one might expect the main effect of semantic fluency across the board (such that reduced semantic fluency might negatively impact both voice and text interfaces alike), the fact that consumers can easily re-read (initially) more difficult to comprehend phrases makes consumers arguably more immune to the negative effect of reduced semantic fluency.

E.7 Study 3

The objectives of Study 3 were twofold. First, the current study was designed to further illuminate the conditions under which voice-based interfaces might enhance versus reduce perceptions of interface flow. Specifically, the current study tests the boundary condition of conversational turns during a voice-based interaction. Second, the current study also measured a set of alternative explanations. Given that flow-like experiences are by definition more positively-valenced experiences, we further assess the possibility that the current findings can be explained merely via a general positive affect account (as opposed to a *specific* positive affective experience, i.e. interface flow). We also further assess whether the mere presence of a voice-based interface might enhance generic attributions of perceived care (given that a voice-based interface actively “talks” to consumers).

E.7.1 Design and Procedure

A total of 610 participants ($M_{\text{Age}}=35.59$, $SD_{\text{Age}}=12.70$, 46.9% males) were recruited from a nation-wide consumer panel (Prolific) and randomly assigned to a 2 (interface modality: voice-based vs text-based interface) $\times$ 2 (conversational turns: high vs. low) between subjects' design. The interface modality conditions mirrored the paradigm used in Study 1. We manipulated the extent of conversational turns building on prior work on turn-taking (Levinson, 2016; Wiemann & Knapp, 1975). Specifically, in the low conversational turns condition, participants answered corresponding questions within one turn (e.g., "Was it a traffic accident, glass breakage, theft or burglary? How did this damage occur? Please describe the course of events as accurate as possible.") whereas in the high conversational turns condition these questions were broken up into separate sub-questions (e.g., Question 1: "Was it a traffic accident, glass breakage, theft, or burglary?", Question 2: "How did this damage occur? Please describe the course of events as accurately as possible."; see Web Appendix A2.3 for further details on the scenario). We predict that separating corresponding questions in a voice- compared to text-based interface interaction will be more detrimental as compared to keeping larger question blocks to consumers' experience of interface flow, given the reduced synchronicity of the interaction (see Table 2). However, increasing the number of conversational turns in the text-based interface might even enhance perceptions of interface flow as text-based communication often unfolds in shorter, more rapid turns (think of the common short message exchanges using social messenger platforms).

Measurement. Mirroring the preceding studies, we used the same items to assess consumers' experience of interface flow ($\alpha_{\text{Flow}} = .91$), experiential service delivery ($\alpha_{\text{ExpS-ervice}} = .94$), and firm evaluation ($\alpha_{\text{FirmEval}} = .94$). To further assess whether the current findings are driven by general positive affect, we measured PANAS (negative items reverse coded; $\alpha_{\text{PosAffect}} = .74$). Finally, we used a single item to assess respondents' feelings of perceived care on a 7-point Likert scale ("How caring did you perceive the interface you used to complete your claim filing procedure?"; see Web Appendix A3.7).

E.7.2 Results

Interface Flow. A two-way ANOVA revealed a significant main effect of the interface condition on interface flow ($F(1, 606) = 13.078$, $p < .001$, $\eta^2 = .02$), demonstrating a significantly higher experience of interface flow in the voice- as opposed to text-based interface modality condition ($M_{\text{VoiceInterface}} = 3.54$, $M_{\text{TextInterface}} = 3.03$, $t = 3.707$, $p < .001$). The conversational turns main effect was non-significant ($F(1, 606) = 0.858$, $p = .35$, η^2

= .001) while the interaction between both experimental factors was significant ($F(1, 606) = 4.206, p < .05, \eta^2 = .007$). Follow-up contrasts with Holm correction for family-wise errors revealed that while a smaller number of conversational turns enhance interface flow in voice-based interfaces, they reduce interface flow in text-based interfaces ($M_{\text{Voice_LowTurns}} = 3.65, M_{\text{Text_LowTurns}} = 2.85, t = 3.995, p < .001$). The presence of more conversational turns in the voice-based interface condition led to a directional reduction of interface flow ($M_{\text{Voice_HighTurns}} = 3.43, M_{\text{Voice_LowTurns}} = 3.65, t = 1.028, p = .47$), and was significant for participants that experienced at least some level of flow ($M_{\text{Voice_HighTurns}} = 4.05, M_{\text{Voice_LowTurns}} = 4.53, t = 2.437, p < .05$; trimmed mean minus half a standard-deviation, see Globerson, 1983).

Service Experience. Mirroring the previous effects on interface flow, a two-way ANOVA assessing the impact on consumers' service experience revealed a significant increase in experiential service delivery in the voice as opposed to text-based interface condition ($F(1, 606) = 37.135, p < .001, \eta^2 = .001$; $M_{\text{VoiceInterface}} = 3.78, M_{\text{TextInterface}} = 2.88$), a non-significant main effect of conversational turns ($F(1, 606) = 0.022, p = .88, \eta^2 = .00$), and a marginally significant interaction between both factors ($F(1, 606) = 3.402, p = .065, \eta^2 = .0007$). Follow-up contrasts with Holm correction confirmed that a smaller number of conversational turns enhanced consumers' service experience in voice-based interfaces, but reduced service experience in text-based interfaces ($M_{\text{Voice_LowTurns}} = 3.97, M_{\text{Text_LowTurns}} = 2.79, t = 5.552, p < .001$). As with flow, the presence of conversational turns in the voice-based interface condition led to a directional reduction of interface flow ($M_{\text{Voice_HighTurns}} = 3.61, M_{\text{Voice_LowTurns}} = 3.97, t = 1.536, p = .25$), and was significant for participants that experienced at least some level of flow ($M_{\text{Voice_HighTurns}} = 4.02, M_{\text{Voice_LowTurns}} = 4.65, t = 2.468, p < .05$; trimmed mean minus half a standard-deviation, see Globerson 1983).

Path Model. Next, we estimated a path model using the lavaan package in R with robust standard errors to assess the overall system of hypotheses. The path model results are summarized in Figure 20. The findings demonstrate that the increase in interface flow evoked by the voice-based interface ($\beta_{\text{Voice}} = .47, z = 3.823, p < .001$), in turn, led to a significant increase in experiential service delivery ($\beta_{\text{Flow}} = .72, z = 28.420, p < .001$) and an overall more positive firm perception in line with our predictions ($\beta_{\text{Flow}} = .45, z = 12.556, p < .001$).

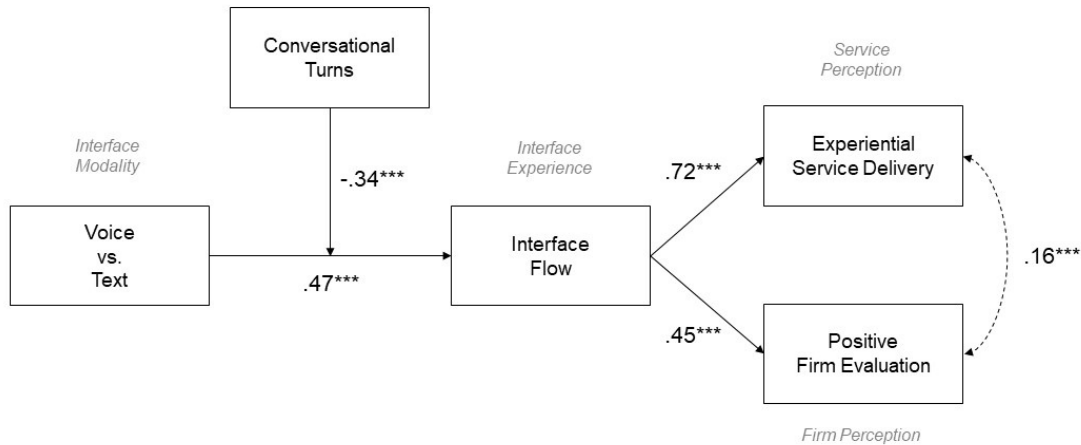


Figure 20. Path model results (Study 3)

Alternative Accounts. Finally, we conducted a series of robustness checks. Specifically, we first tested whether interface modality might enhance general positive affect. Given that flow-like experiences are a specific type of positive affect, it is conceivable that voice-based interfaces are not specific to interface flow but drive general positive affect more broadly. Ruling out a general positive affect account, we found no main effect of modality on positive affect ($F(1, 606) = 1.121, p = .29, \eta^2 = .00$; $M_{\text{VoiceInterface}} = 4.64, M_{\text{TextInterface}} = 4.53$) and no interaction between both factors ($F(1, 606) = 1.834, p = .18$). Finally, we also found no effect of the interface modality on perceived care ($F(1, 606) = 0.086, p = .77, \eta^2 = .00$; $M_{\text{VoiceInterface}} = 3.35, M_{\text{TextInterface}} = 3.31$; also no interaction between both factors, $F(1, 606) = 2.014, p = .16$).

E.7.3 Discussion

The results of Study 3 demonstrate that a greater number of conversational turns can reduce interface flow when using voice-based as opposed to text-based interfaces. While one might expect the main effect of conversational turns across the board (such that enhancing the number of conversational turns might negatively impact both voice-based and text-based interfaces alike), the high synchronicity and faster pace of voice-based interactions makes consumers more sensitive to the negative effect of an enhanced number of conversational turns. The current study also rules out that these findings can be explained merely by a general positive affect account or by enhancing attributions of perceived care.

E.8 Study 4

The objectives of Study 4 were twofold. First, the current study was designed to examine whether the current effects are specific to consumers' experience of interface flow.

Flow-like experiences have been shown to be driven by creating “optimal levels of challenge”, i.e. experiences that are neither considered as boredom nor as excessively challenging. Thus, for flow to emerge, there needs to be some level of challenge (see optimal stimulation theory; Holbrook & Gardner, 1993) consistent with the faster paced, more synchronous, and more flexible cue and channel characteristics of voice-based interactions (see Table 2). The current study tests this flow-specific mechanism. Second, the current study is also designed to further examine the impact of voice-based interfaces on a range of behavioral outcomes.

E.8.1 Design and Procedure

All methods and statistical analyses for this study were pre-registered (https://aspre-dicted.org/NZH_966). A total of 811 participants ($M_{\text{Age}}=35.15$, $SD_{\text{Age}}=12.93$, 50.92% males) were recruited from Prolific Academic and randomly assigned to either a voice-based or a text-based interface. The interface conditions used the same insurance context as in all preceding studies. The first part of the task was identical to all three preceding studies (filing a claim).

We created a MailChimp landing page for our fictitious insurance brand TER Insurance and developed and designed an advertisement that provided the user with an option to sign up as a future user, as a consequential, behavioral measure of user conversion. The advertisement was presented at the end of the study to each participant with a promotion to become a beta user to test future insurance services (see Web Appendix A4. for the advertising stimuli). In both conditions the advertisement was identical, except that in the voice condition the advertisement asked the user to sign-up to “become a test user for our new insurance services based on our *voice* interface” while in the text-based interface condition, the advertisement asked the user to sign up to “become a test user for our new insurance services based on our *text* interface”. The main dependent variable was the conversion rate of the advertisements and measured as the number of participants subscribing divided by the number of unique page visits.

Measurement. The same items as in the preceding studies were used to assess consumers’ experience of interface flow ($\alpha_{\text{Flow}} = .91$), experiential service delivery ($\alpha_{\text{ExpService}} = .92$), and their perception of the service firm ($\alpha_{\text{FirmEval}} = .94$). To assess the flow-specific mechanism for consumers’ experienced challenge, we used a 6-item scale (adapted by Novak et al., 2000, $\alpha_{\text{Challenge}} = .89$; see Web Appendix A3.8). To assess consumer sentiment, we used a rating task inspired by Hildebrand & Bergner (2020) using an open-response technique about consumers’ positive and negative thoughts and feelings during

the service task. We used the tidytext package in R to process the text data and used the sentimentr package to extract sentiment scores (using both positively and negatively valenced words) (Nielsen, 2011). Finally, we measured the subscriber conversion following McDowell et al. (2016) (subscribers divided by the number of page visits).

E.8.2 Results

Interface Flow & Service and Firm Perception. In line with our theorizing and corroborating the results of our previous studies, participants who used the voice-based interface experienced significantly higher levels of interface flow compared to participants using the text-based interface ($M_{\text{VoiceInterface}} = 4.02$, $M_{\text{TextInterface}} = 3.25$, $t = 6.96$, $p < .001$; Cohen's $d = .49$). Mirroring the results of the preceding studies, we also found that the voice-based interface led to a significantly more experiential service delivery than the text-based interface ($M_{\text{VoiceInterface}} = 4.23$, $M_{\text{TextInterface}} = 3.01$, $t = 11.25$, $p < .001$; Cohen's $d = .79$) and a more positive evaluation of the service firm ($M_{\text{VoiceInterface}} = 4.75$, $M_{\text{TextInterface}} = 4.50$, $t = 2.26$, $p < .05$; Cohen's $d = .16$).

Flow & Optimal Challenge. To test the flow-specific mechanism, we estimated a mediation model (5,000 bootstrap samples) with the interface modality condition as the independent variable, consumers' experienced challenge as the mediator, and interface flow as the dependent variable. As expected, the voice-based interface led to a significant increase in consumers' experienced challenge ($\beta_{\text{Interface}} = .32$, $SE = .09$, $t = 3.47$, $p < .001$). In turn, greater levels of experienced challenge led to a higher flow perception ($\beta_{\text{Challenge}} = .51$, $SE = .03$, $t = 13.03$, $p < .001$), and a significant indirect effect with a confidence interval excluding zero ($\beta_{\text{Indirect}} = .16$, 95% CI: [.07, .25]), indicating full mediation.

To further assess whether more excessive levels of challenge negatively impact interface flow (as a test of the "optimal level of challenge" notion of flow), we estimated a polynomial regression in which we predicted interface flow using the interface modality condition and the second-degree polynomial of experienced challenge. The model was significant ($F(3, 807) = 78.14$, $p < .001$), confirming that the voice-based interface led to a significant increase of flow ($\beta_{\text{Interface}} = .60$, $SE = .10$, $t = 5.96$, $p < .001$), a significant positive main effect of experienced challenge ($\beta_{\text{Challenge}} = 18.78$, $SE = 1.44$, $t = 13.07$, $p < .001$), and also a significant negative squared effect of challenge ($\beta_{\text{Challenge}^2} = -3.18$, $SE = 1.43$, $t = -2.23$, $p < .05$), conforming the expected "optimal level of challenge" necessary for flow-like user experiences (i.e., reduced flow at extreme levels of challenge).

Behavioral Outcomes. We observed a systematic increase in consumers' behavioral responses across our key dependent variables (i.e., consumer sentiment, conversation rate). First, analyzing consumer sentiment scores revealed a significant effect of interface type on sentiment valence ($F(1, 809) = 23.73, p < .001$). Participants who interacted with the voice-based interface perceived the service experience significantly more positive than those who interacted with the text-based interface ($M_{VoiceInterface} = .17, SD = .41$; $M_{TextInterface} = .03, SD = .42$). For example, while participants in the voice-based interface condition mentioned that the experience was “exciting”, “fun”, and “like speaking to a real customer service representative”, while participants in the text-based interface only mentioned how “efficient”, “simple”, and “practical” the claim interface was. Next, a path model using the lavaan package in R with robust standard errors demonstrates that the increase in interface flow evoked by the voice-based interface ($\beta_{VoiceInterface} = .77, z = 6.92, p < .001$) led to a significant increase in experiential service delivery ($\beta_{Flow} = .76, z = 36.81, p < .001$) and an overall more positive firm evaluation ($\beta_{Flow} = .66, z = 25.69, p < .001$), which in turn significantly enhanced consumer sentiment ($\beta_{ExpService} = .03, z = 2.52, p < .05$; $\beta_{FirmEval} = .09, z = 8.76, p < .001$). Finally, a Chi-Square test assessing the subscription rate conditional on the type of interface further confirmed that participants in the voice-based interface condition were also significantly more likely to subscribe to become a future user of TER insurance compared to participants in the text-based interface condition ($M_{VoiceInterface} = 68.37\%$, $M_{TextInterface} = 50.05\%$; $\chi^2(1, N = 201) = 5.93, p = .01$; a path model analysis was not possible at the individual level due to the inability to track participants between systems).

E.8.3 Discussion

Study 4 demonstrates that consumers are more likely to engage in positive firm advocacy (i.e., enhanced consumer sentiment about the firm and conversion rates) when using voice-based compared to text-based interfaces. The current study provides also a more nuanced look at the underlying mechanism of flow, demonstrating that the increase in interface flow is driven by an increase in consumers' experienced level of challenge. We also provide initial evidence consistent with an “optimal challenge” account of flow-like experiences such that the effect of interface flow is weakened when consumers experience the service interaction as extremely challenging (weakened but not turned-off).

E.9 General Discussion

Across four studies, we shed light on how the affordances of voice-based interfaces impact consumers' service experiences. We demonstrate that voice-based compared to

text-based interfaces cause more absorptive, flow-like service experiences. In addition, we provide evidence for two novel boundary conditions (semantic fluency and conversational turns) that can reduce flow-like experiences and in turn dampen the observed positive effects of voice-based interfaces. Finally, we show how an increase in interface flow impacts both firm and service perceptions (how consumers perceive the overall service experience and evaluate the firm) as well as behavioral outcomes (contract renewal, consumer sentiment, conversion rate; see Table 23 for a summary of results and key findings across studies).

Table 23. Summary of studies

Study	Condition	Interface Flow		Exp. Serv. Delivery		Firm Evaluation		Contract Renewal		Consumer Sentiment		Conversion Rate %	Sample Size	Key Findings
		M	SD	M	SD	M	SD	%	%	M	SD			
1	Text	2.31	1.67	2.90	1.92								184	Voice-based (as opposed to text-based) interfaces lead to enhanced levels of interface flow. Enhanced levels of interface flow in turn increase experiential service perceptions.
	Voice	3.09	1.84	3.78	1.73									
2	Text – Low Fluency	2.86	1.57	2.65	1.57	4.21	1.72	64.24%					612	Semantic fluency moderates the effect of interface modality on interface flow such that the positive effect of voice-based interfaces on interface flow is enhanced (reduced) under conditions of high (low) fluency. Voice-based (as opposed to text-based) interfaces enhance attributions of experiential service experience and firm evaluation.
	Text – High Fluency	2.59	1.47	2.39	1.38	4.67	1.45	69.18%						
	Voice – Low Fluency	3.38	1.63	3.40	1.67	4.58	1.64	63.92%						
	Voice – High Fluency	3.77	1.61	3.76	1.46	5.16	1.47	64.34%						
3	Text – Low Turns	2.85	1.63	2.79	1.76	4.98	1.43						610	The extent of conversational turns moderates the effect of interface modality on interface flow such that the positive effect of voice-based interfaces is reduced with a greater number of conversational turns.
	Text – High Turns	3.20	1.69	2.98	1.78	5.10	1.48							
	Voice – Low Turns	3.65	1.85	3.97	1.80	4.95	1.65							
	Voice – High Turns	3.43	1.58	3.61	1.74	5.20	1.42							
4	Text	3.25	1.52	3.01	1.51	4.50	1.54			.03	.42	50.05%	811	Voice-based (as opposed to text-based) interfaces enhance positive firm-related behaviors such as more positive consumer sentiment and an increase in conversion rates.
	Voice	4.02	1.62	4.23	1.58	4.75	1.61			.17	.41	68.37%		

E.9.1 Theoretical implications

The findings of this research make four novel contributions. First, to the best of our knowledge, the current research is one of the first to contrast the effects of voice-based versus text-based self-service technologies on perceptual and behavioral service outcomes. We provide a novel theory of voice-mediated service interactions, hypothesizing and empirically demonstrating that the sensory-rich experience of voice-based interfaces leads to more positive service perceptions and an increase in interface flow. Specifically, we provide a novel look at prior work on flow (A. Berger et al., 2018; Hoffman & Novak, 1996; Zanjani et al., 2016), demonstrating that the voice modality is a key driver of flow and in turn enhances both perceptual (consumers' perception of the firm and service delivery) as well as behavioral service outcomes (from contract renewal to conversion rates).

Second, the current findings offer novel insights into consumer-technology relationships (Hoffman & Novak, 2018; Novak & Hoffman, 2018) and how the natural interaction through a voice-based interface creates a new “assemblage” of consumer experiences. Specifically, this research provides evidence that even though speaking to a voice-based interface represents merely a change in communication modality, consumers tend to form experiences that impact not only the assemblage itself (i.e., the relationship between the consumer and the interface) but also the entity providing the interface experience (i.e., the firm) by both attributing more positively-valenced perceptions toward the firm and to extend the relationship with that firm in the future.

Third, the current work introduces and empirically tests two novel boundary conditions that can either enhance or dampen the positive effects of voice-based user interfaces. First, we demonstrate that the use of easy-to-process phrases (i.e., high semantic fluency) is critical to boost greater levels of interface flow and in turn positively impact both service and firm perceptions, as well as behavioral outcomes. Second, we also show that consumers' experience of flow decreases with an increasing number of conversational turns throughout the conversation due to the more disjointed flow of the conversation. These findings provide a new look at the conditions under which voice-based interfaces can either enhance or reduce flow-like service experiences and the critical importance of the “conversational design” of voice-based interfaces.

Finally, the current findings also contribute to prior work on experiential service delivery. Experiential service delivery has been often considered to depend upon a multi-

sensory service design (de Oliveira Santini et al., 2020), one that often requires a systematic orchestration of different service and context variables (Zomerdijk & Voss, 2010). However, the current findings demonstrate that a more experiential service perception can be induced by merely altering the interface modality without changing any other aspect of the service experience. The results of the current work suggest that voice-based interactions can transform even mundane tasks such as filing an insurance claim into an absorptive and ultimately more satisfactory service experience for consumers.

E.9.2 Managerial implications

The findings of this research have immediate implications for practitioners. In what follows, we provide an overview the implications and design guidelines for industry practice.

Simple language and conversational flows. The semantic and structural characteristics of voice-based interfaces greatly impact the user experience and firm perception. Practitioners should carefully test and ensure that the conversational design of their service interaction is adequate given the user and task (i.e., ensuring high semantic fluency). In short, the boundary conditions we examined in this study highlight that the conversational design of a voice-based interface is sensitive to both the number of conversational turns during a service interaction and the specific words used by a voice interface during the conversation. When implementing voice-based interfaces, practitioners should be careful in limiting the number of conversational turns required to complete the service, as well as optimizing semantic fluency by using simpler and more familiar words to allow easier processing and in turn an enhanced service experience.

More inclusive user experiences. Voice-based interfaces not only provide a richer, more absorptive user experience as demonstrated by this research but are also more inclusive across consumer demographics. Specifically, consumers with impaired writing abilities such as dyslexia or impaired sight (such as older consumers or consumers with a poor vision) can directly interact with a voice-based interface across service settings. Traditional interfaces often lead to detrimental outcomes and heightened frustration particularly for those otherwise disadvantaged consumer segments (Abdolrahmani et al., 2018). Thus, the well-selected use of voice technology provides an unexplored potential to deliver not only more sensory-rich but ultimately more inclusive user experiences.

Voice-based interfaces as a sales channel. The findings of this research revealed that voice-based interfaces can boost downstream firm-related outcomes such as conversion or contract renewal rates compared to text-based interfaces. It is plausible that this

greater persuasiveness is driven by a combination of both enhanced flow (as per our theorizing) and also mechanisms studied in earlier work (such as enhanced social presence; Qiu & Benbasat, 2009). The important implication for firms is that voice-based interfaces are an effective lever in a marketer's sales automation toolbox. However, our boundary conditions also reveal that firms are well-advised to ensure that consumers' processing capacities are not overwhelmed (i.e., by reducing the complexity of the language used and by avoiding a large number of conversational turns). Thus, firms need to balance the effectiveness of voice-based interfaces for sales automation purposes with the affordances of the task and the user.

Context-dependent channel choice. Current industry practice in choosing the right channel to interact with consumers is often more driven by trial and error and technology vendors as opposed to a strategic mapping of task affordances and business objectives. We hope that the current research can provide guidelines to decide more effectively under which conditions a voice-based interface is appropriate and when not (see also Table 22 for a summary of key properties of voice-based communication). Given the high channel synchronicity of voice-based interfaces, service firms are advised to choose text-based interfaces for tasks that require more deliberation from the consumer and that are better suited for asynchronous interactions. Specifically, service tasks that require parallel processing such as comparing multiple product configurations during a shopping task are better suited for text-based interfaces as the consumer can re-read and re-type information if needed.

Monitoring channel choice. Multimodal interfaces are becoming increasingly common across the service value chain and allow consumers to interact with self-service technologies via voice or text interchangeably. This begs the question of whether consumers should be able to choose their preferred interaction modality or not. Our results suggest that some consumers may feel overwhelmed by the speed and immediacy of voice-based interfaces (Study 4). Hence, firms are advised to carefully monitor drop-out rates and also transition periods when consumers switch from one modality over to the other. Allowing consumers to choose their preferred channel is a critical ingredient in enhancing consumers' agency during service interactions which we expect to positively impact both service and firm outcomes.

Cost-effective integration in existing infrastructure. With the rapid development of technology platforms such as Google's natural language processing platform Google Dialogue or Amazon's Polly, voice-based interfaces can be easily integrated into the cor-

porate infrastructure. Instead of developing a native and firm-owned voice-based interface channel, companies can easily tweak existing language models (such as Google WaveNet as in the current research) and incorporate them into their current service delivery process. As highlighted in our technical documentation in the Web Appendix (see Web Appendix A1), the deployment and technical integration into existing enterprise solutions are both cost-effective (infrastructure provided, maintained, and continuously improved by firms like Google or Amazon) and customizable (such as developing a unique “voice character” that matches the personality of the brand). In short, the use of voice-based interfaces is highly modular and can be integrated flexibly across different stages of the customer journey and without requiring substantive changes to the corporate technology infrastructure.

E.9.3 Future Research

We see three immediate directions for future research. First, future work could further explore the effective “voice design” and to which extent the voice of the assistant or interface (such as the depth, pitch, or other vocal features of the voice-based interface) can be mapped onto the “brand personality” that the voice-based interface represents. Second, one unexplored area of research is whether and to which extent the specific features of a voice-based interface can evoke specific affective responses from the user interacting with the interface. As softer and slower-paced voices tend to induce greater psychological comfort and calmness (Elkins & Derrick, 2013; Nass & Lee, 2001), future work might explore which specific features of the voice-based interface affect which specific psychological user experience. Third, the use of voice technology may not be beneficial across all settings or contexts (Cambre et al., 2020; Seaborn et al., 2021). We acknowledge that the impact of voice-based interfaces on consumers is arguably multiply determined, which is also reflected in the varying effect sizes across our studies. Future work could therefore further explore under which conditions voice technology can enhance consumers’ experience versus when it becomes a nuisance or even detrimental.

E.10 Web Appendix

E.10.1 A. Supplementary Methods

A1. Technical Details Interface Implementation

A1.1 System Architecture Voice Interface

The voice assistant application was built using the open-source Python framework "Flask" for the backend and the open-source JavaScript framework "VueJS" for the frontend. They act as separate components with the frontend running in the user's browser and interacting with the backend API via JSON.

A custom JSON schema was devised to configure the user journeys which enables simple branching and looping algorithms based on configurable regex checks. Thus, we can create adaptive user journeys based on user's verbally issued intents. The frontend is voice-based and requires a user to give the permission to access the microphone. Voice input is transcribed via the Google Speech API. All data is stored in an SQLite database. If desired, the voice files can be stored in the server's file system.

A1.2 Details on Voice Interface Configuration

In order to illustrate the functionalities of our interface in more detail, we discuss here our JSON File that defines how a user journey is configured. Each journey configures the *id*, *title*, and *description*, based on which different journeys can be called in parallel. Note that the *description* can contain HTML parts, which can be used to individualize the frontend of the interface.

```
{
  "id": "JOURNEY_ID",
  "title": "My Voice Assistant",
  "description": "<strong>Hi dear user!</strong><br>Welcome to the voice
assistant.",
  "waypoints": {
    // ...
  }
}
```

To illustrate the operational functionality of our interface, we showcase two types of journeys that can be built with our interface. The first is a simple linear user journey. The actual journey is defined with the waypoints key. The different steps are connected with simple links: each link has a specific type (in this case a GO_TO) that defines a link_to value that points to the key of the waypoint that the user should continue with after the given step. Each step contains a predefined text and/or audio file. As a result, simple linear user journeys can be modelled as follows:

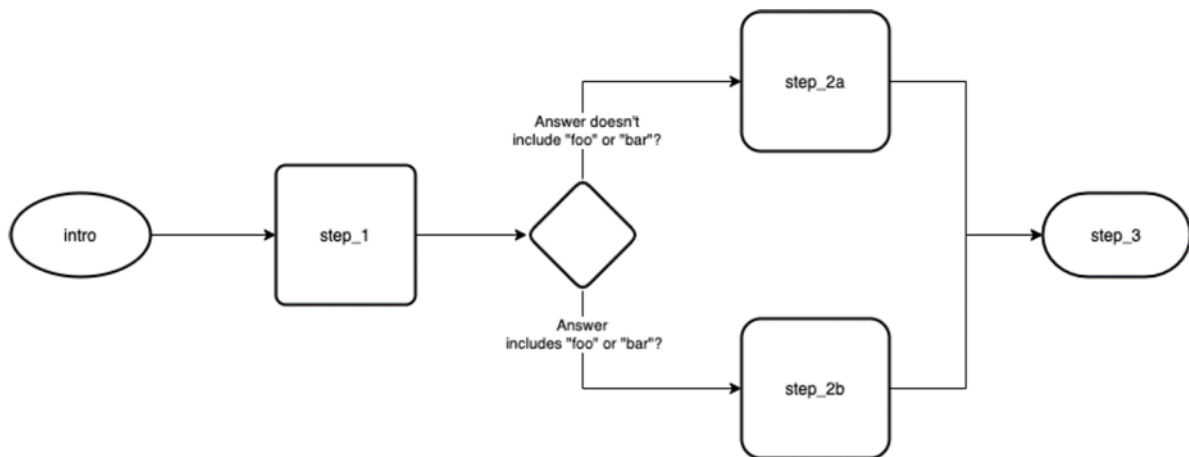


```

{
  "waypoints": {
    "intro": {
      "text": "Hi there, how are you?",
      "audios": {
        "audio_default": "/static/audios/simple/intro.mp3"
      },
      "next": {
        "type": "GO_TO",
        "data": {
          "link_to": "step_1"
        }
      }
    },
    "step_1": {
      "text": "Wonderful, I'm doing great thank you!",
      "audios": {
        "audio_default": "/static/audios/simple/step_1.mp3"
      },
      "next": {
        "type": "GO_TO",
        "data": {
          "link_to": "step_2"
        }
      }
    },
    "step_2": {
      "text": "Thanks a lot, bye!",
      "audios": {
        "audio_default": "/static/audios/simple/step_2.mp3"
      }
    }
  }
}

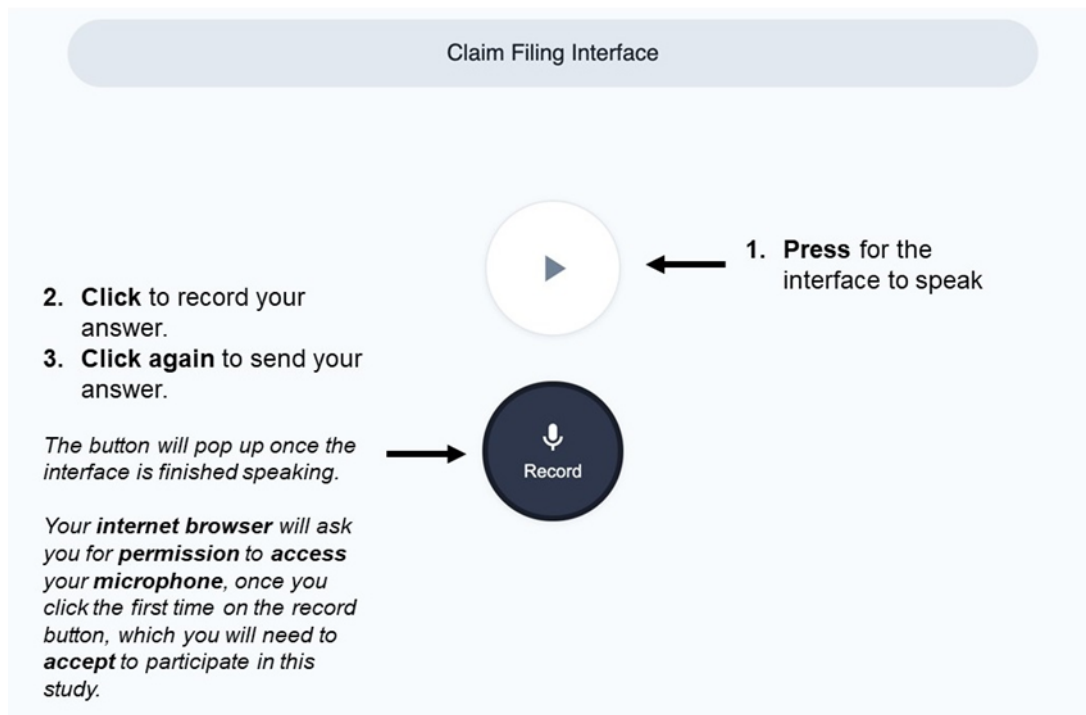
```

Second, in order to create adaptive user journeys, we can use a so-called Simple_Branch. As we kept the conversation flow steady in the current study, we did not use this link type. However, to demonstrate the potential of our interface for modelling more complex, adaptive conversation flows, the below illustration highlights the option to include additional branching (if needed for future research implementations for instance):



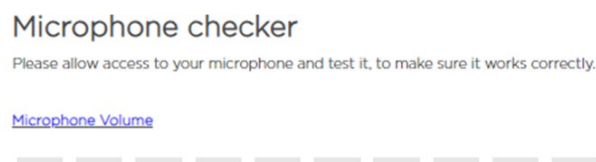
```
{
  "next": {
    "type": "SIMPLE_BRANCH",
    "data": {
      "options": [
        {
          "match": "^(?!.*(foo|bar)).*$",
          "link_to": "step-1a"
        },
        {
          "link_to": "step-1b"
        }
      ]
    }
  }
}
```

A1.3 Screenshot Voice Interface



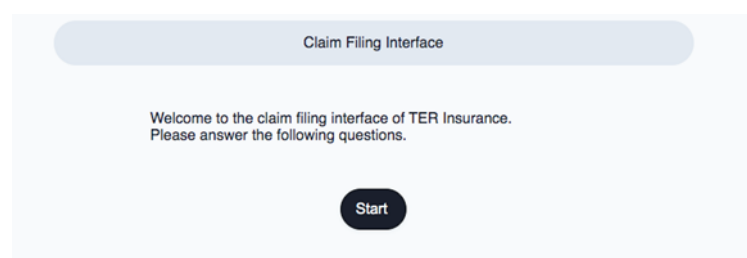
A1.4 Screenshot Voice Interface

Prior to initiating our experiment, we presented participants in the voice conditions with a microphone check. Participants were instructed to provide access to their microphone input and test it. Only participants for whom the interface detected a voice input were allowed to proceed with the study.

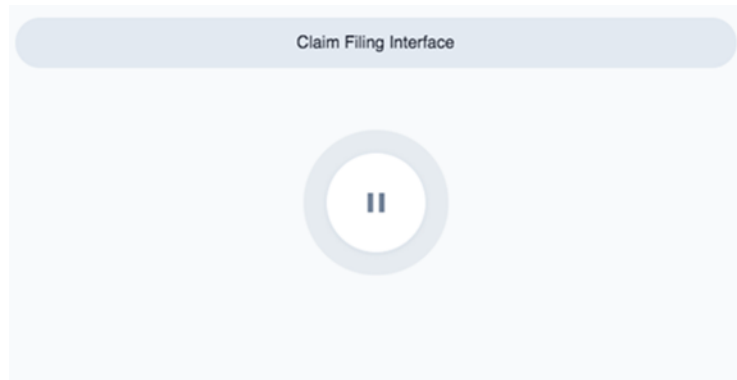


A1.4 Exemplary User Journey Voice Interface

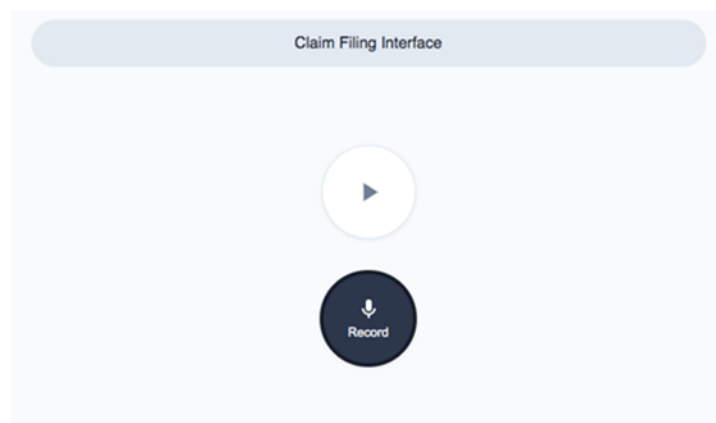
The user interaction with the voice-interface follows a simple conversational path, as illustrated below:



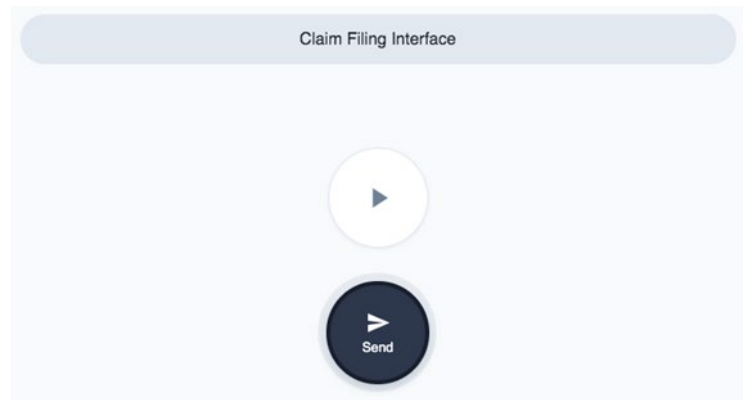
- 1) “Start”-screen: This is the first screen the user interacts with. The conversation flow is initialized by clicking on the “start”-button.



- 2) “Listening”-screen: The conversational interface plays its message. Concentric vibrating circles graphically underline the audio message.



- 3) “Interaction”-screen: The user can either play the audio message again by clicking on the “play”-button (based on which the “listening-screen” will arise) or to record a message by clicking on the “record” button.



4) “Record”-screen: Once the user clicks the “record”-button, the following screen appears. The user can now record a message by speaking into the microphone. When the user has finished the recording, the message can be sent by clicking on the “Send”-button. Either the user is linked to the next step of the predefined interaction journey (“interaction”- screen) or directly transferred to the survey platform once the interaction is finished.

A1.6 Exemplary Text-Based Interface

Claim Filing Interface

What is the reason for your claim? Was it (1) a traffic accident, (2) a glass breakage, (3) a theft or burglary, (4) a damage caused by a natural event, or (5) another incident?

CONTINUE

Claim Filing Interface

How high do you estimate the damage?

CONTINUE

Sample screenshots from the insurance claim filing task with the text-based interface.

A2. Insurance Claim Filing Task

A2.1 Main Task

Instructions:

Please imagine the following scenario:

- Five days ago, you drove to the city centre, parking your car in a public parking garage. On the way out of your parking spot, you accidentally hit a pole, leaving a long scratch on the left side of your car. Eventually, you visited the car repair yesterday and you were told that the paintwork must be completely redone. The damage is estimated to be around \$1250.
- You now want to file a claim concerning your car damage with your car insurance company TER Insurance.
- On the following page(s), you are asked to complete your claim via a web interface, which is automatically sent to your insurance company. After responding to all questions, please continue with the rest of the survey.

Follow-up questions as part of the Claim Filing Process:

- “What happened? Was it a (1) traffic accident, (2) glass breakage, (3) theft or burglary, (4) damage caused by natural event, (5) another incident?”
- “How did the damage occur? Please describe the course of events as accurate as possible.”
- “What damage occurred?”
- “How high do you estimate the damage?”
- “When did you notice the damage?”
- “Would you like to notify us about something else?”

Each follow-up question was presented on a separate page, and responses were given in open text fields for the text-based interface conditions.

Participants in the voice-based interface conditions were provided with exactly the same questions, the same wording, and the same sequence. They had to click a button to listen to each follow-up question and another button to record their voice input for their responses to each question (see voice interface screenshots in section A1).

A2.2 Variations for Study 2

- Added additional follow-up questions to the claim filing process:
 - “What repairs and maintenance do you think are needed to fix the damage?”
 - “Next, please tell us about any other parties that might be involved in this case by addressing the following questions. Were there any other directly or indirectly involved parties in the accident and if so which?”
 - “Additionally, please notify us if you already consulted a technical expert on the damage?”
 - “Finally, please indicate if you have been already in touch with another insurance about this damage.”
- Added another interaction task (“follow up customer questionnaire”), where we asked participants to indicate where they want to continue their policy:
 - “We just saw that your insurance policy is about to end. Would you like us to send you an email so that you can renew your policy: yes or no?”
(*low semantic fluency*)
 - “It has been determined that your insurance policy is due to expire shortly. Please indicate if you would like to receive an email to prolong your policy: yes or no?” (*high semantic fluency*)

A2.2 Variations for Study 3

We manipulated the number of conversational turns through coupling (decoupling) of the follow-up questions in individual interaction turns. In the low conversational turns conditions, the questions were combined into 4 individual interaction turns. In the high conversational turns condition, each question asked in claim filing process resembles an own interaction turn resulting in a total of 10 interaction turns.

A3. List of Scales

A3.1 Interface Flow (Study 1, Study 2, Study 3, Study 4)

“To what extent do you agree with the following statements?” (7-point Likert scale; “Strongly disagree” to “Strongly agree”)

- i. This interface totally absorbed me.
- ii. This interface really intrigued me.
- iii. This interface drew me in.

Adapted from Unger & Kernan (1983)

A3.2 Experiential Service Delivery (Study 1, Study 2, Study 3, Study 4)

“How do you feel about filing your insurance claims through this interface?” (7-point Likert scale; “Strongly disagree” to “Strongly agree”)

- i. It offers novel experiences.
- ii. It satisfies my sense of curiosity.
- iii. There is novelty in it.
- iv. I feel like I am exploring new worlds.

Adapted from Unger & Kernan (1983)

A3.3 Service Firm Evaluation (Study 2, Study 3, Study 4)

“To what extent do you agree with the following statements?” (7-point Likert scale; “Strongly disagree” to “Strongly agree”)

- i. Overall, I am satisfied with the online experience with TER Insurance.
- ii. My overall opinion of the services provided by TER Insurance is very good.

Adapted from G. G. Lee & Lin (2005).

A3.4 Insurance Familiarity Questionnaire (Study 1, Study 2, Study 3, Study 4)

In order to better understand or control for participants’ insurance claim filing experience, we asked the following items:

Car Ownership: “Do you own a car, or have you ever owned a car?”

- Yes
- No

Possession of an Insurance: “Do you currently own a car insurance?”

- Yes
- No

Previous Experience with Insurance Claim Filing: “Have you ever filed an insurance claim (in person, via phone, via the insurance company’s website)?”

- Yes
- No

When participants replied with “Yes” to the previous experience question, they were asked the following item:

Last File Claiming Experience: “When have you last filed an insurance claim?”

- Less than a year ago
- Between a year and two years ago
- Between two and three years ago
- More than three years ago

A3.5 Experience with Voice-Based Interfaces (Study 1, Study 2, Study 3, Study 4)

Previous Experience with Voice Assistants: “Have you ever used a voice assistant, such as Siri (Apple), Google Assistant (Google), or Alexa (Amazon)?”

- Yes
- No
- I don’t know

Voice Assistant Use: “How often do you use voice assistants?”

- Never
- Less than once a month
- Once a month
- Once every two to three weeks
- Once a week
- Several times a week
- Once a day
- Several times a day

Level of Comfort with Voice Assistant Use: “How comfortable are you using a voice assistant?”

- Not at all
- Slightly
- Somewhat
- Moderately
- Extremely

A3.6 Readability Scores (Study 2)

Semantic fluency was assessed using three different readability scores: the Flesch reading ease score (FRE), the Automated Readability Index (ARI), and the New Dale-Chall readability formula (NDR). The formulas to the individual scores are presented below:

$$FRE = 206.835 - 1.015 \times \left(\frac{\text{words}}{\text{sentences}} \right) - 84.6 \times \left(\frac{\text{syllables}}{\text{words}} \right)$$

$$ARI = 4.71 \times \left(\frac{\text{characters}}{\text{words}} \right) + 0.5 \times \left(\frac{\text{words}}{\text{sentences}} \right) - 21.43$$

$$NDR = 0.1579 * (\text{Percentag of difficult words}) + 0.0496 \\ * \text{Average Sentence Lengh in Words}$$

A3.7 Perceived Care of the Interface (Study 3)

“To what extent do you agree with the following statement?” (7-point Likert scale; “Strongly disagree” to “Strongly agree”)

- i. I feel like the claim filing interface actually cares about me

Adapted from Breivik & Thorbjørnsen (2008).

A3.8 Challenge (Study 4)

“To what extent do you agree with the following statement?” (7-point Likert scale; “Strongly disagree” to “Strongly agree”):

- i. Using the claim filing interface challenges me.
- ii. Using the claim filing interface challenges me to perform the best of my ability.
- iii. Using the claim filing interface provides a good test of my skills.
- iv. I find that using the claim filing interface stretches my capabilities to my limits.

“Please answer the following questions:” (7-point Likert scale; “Not at all challenged” to “Extremely challenged”):

- v. How much does using the claim filing interface challenge you, compared to other things you do on the computer?
- vi. How much does using the claim filing interface challenge you, compared to the sport or game you are best at?

Adapted from Novak et al. (2000)

A4. Advertising Stimuli

A4.1 Voice-Based Interface Advertising Stimuli

TER Insurance

Our Digital Insurance

Incident



↓

Claim with Voice Interface



↓

Solution



Leave your email and become a test user for our new insurance services based on our voice interface!

Super fast claim resolution. Fair prices. Insurance made easy.

Email Address

Yes, I want to become a test user.

A4.2 Text-Based Interface Advertising Stimuli

TER Insurance

Our Digital Insurance

Incident



↓

Claim with Text Interface



↓

Solution



Leave your email and become a test user for our new insurance services based on our text interface!

Super fast claim resolution. Fair prices. Insurance made easy.

Email Address

Yes, I want to become a test user.

E.10.2 B. Supplementary Analyses

B1 Demographics Across Studies

	N	N male	N fe- male	N other	Mean age	SD age	% male
Study 1	184	92	90	2	32.97	11.05	50.0%
Study 2	612	278	322	12	37.19	13.34	45.4%
Study 3	610	286	313	11	35.59	12.70	46.9%
Study 4	811	392	413	6	35.15	12.93	48.3%

B2 Demographics Across Studies

<i>Predictors</i>	Model 1				Model 2				Model 3				Model 4			
	<i>Est.</i>	<i>CI</i>	<i>p</i>	DV: Flow	<i>Est.</i>	<i>CI</i>	<i>p</i>	DV: Flow	<i>Est.</i>	<i>CI</i>	<i>p</i>	DV: Flow	<i>Est.</i>	<i>CI</i>	<i>p</i>	DV: Flow
(Intercept)	2.31	1.95- 2.57	<.001		2.10	1.66- 2.53	<.001		2.08	1.58- 2.57	<.001		1.68	0.58- 2.78	.003	
Interface Modality (Voice)	0.78	0.27- 1.29	.003		0.77	0.26- 1.28	.003		0.81	0.09- 1.53	.028		1.88	0.48- 3.28	.009	
Gender (Male)					0.44	-0.07- 0.95	.089		0.48	-0.23- 1.19	.184		0.50	-0.22- 1.21	.172	
Interface Modality x Gender									-0.9	-1.10- 0.93	.869		-0.18	-1.21- 0.58	.735	
Experience with VAs (No)													-0.88	-3.63- 1.87	.529	
Experience with VAs (Yes)													-1.14	-2.43- 0.15	.083	
Car Ownership (Yes)													0.23	-0.64- 1.09	.607	
Domain Experience													0.10	-0.48- 0.69	.727	
Observations	184				184				184				184			
R2 / R2 adjusted	.047/.042				.062/.052				.062/.047				.083/.047			

B3 Discriminant Validity Tests Interface Flow vs. Experiential Service Experience

Our theorizing rests on the assumption that interface flow and experiential service experience are two conceptually distinct constructs. To rule out that interface flow and experiential service experience are reflecting the same underlying construct, we conducted a series of discriminant validity tests.

We first estimated and tested a two-factor model for each construct against a single-factor model with all items loading on the same factor to assess the discriminant validity between the constructs (Bagozzi et al., 1991). A significant Chi-Square Difference Test indicates the superiority of the two-factor over the single-factor model across all studies (Study 1: $\chi^2_{\Delta}(1) = 231.49, p < .001$; Study 2: $\chi^2_{\Delta}(1) = 694.32, p < .001$; Study 3: $\chi^2_{\Delta}(1) = 384.10, p < .001$; Study 4: $\chi^2_{\Delta}(1) = 379.2, p < .001$), and smaller information criteria in favor of the two-factor model across each study (Study 1: $BIC_{TwoFactor} = 4108.6$ vs. $BIC_{SingleFactor} = 4334.9$; Study 2: $BIC_{TwoFactor} = 16235$ vs. $BIC_{SingleFactor} = 16923$; Study 3: $BIC_{TwoFactor} = 17726$ vs. $BIC_{SingleFactor} = 18103$; Study 4: $BIC_{TwoFactor} = 17965$ vs. $BIC_{SingleFactor} = 18337$). Second, following Fornell and Larcker (Fornell & Larcker, 1981), discriminant validity was also supported when comparing the constructs' average variance extracted (AVE) with the squared correlation of the two constructs across all studies (Study 1: $AVE_{Flow} = .85, AVE_{Experiential} = .81; r^2 = .56$; Study 2: $AVE_{Flow} = .77, AVE_{Experiential} = .79; r^2 = .50$; Study 3: $AVE_{Flow} = .77, AVE_{Experiential} = .75; r^2 = .60$; Study 4: $AVE_{Flow} = .78, AVE_{Experiential} = .75; r^2 = .61$). Third, estimating the bootstrapped correlation coefficients between the two constructs (5000 iterations; using the 'boot' package in R) across studies confirm that the confidence intervals do not include either 1 or -1 (Study 1: $CI_{95\%} = [.68; .83]$; Study 2: $CI_{95\%} = [.66; .75]$; Study 3: $CI_{95\%} = [.74; .80]$, Study 4: $CI_{95\%} = [.75; .81]$). These results further indicate that the two constructs are distinct (Bagozzi et al., 1991).

Overall, the discriminant validity tests demonstrate that experiential service experience and interface flow are distinct and statistically separable constructs.

Paper F – “I Will Follow You!” – How Recommendation Modality Impacts Processing Fluency and Purchase Intention

Table 24. Bibliographical metadata Paper F

Title	“I Will Follow You!” – How Recommendation Modality Impacts Processing Fluency and Purchase Intention
Authors	Schwede, Melanie; Zierau, Naim; Janson, Andreas; Hammerschmidt, Maik; Leimeister, Jan Marco
Publication outlet	Proceedings of the 33rd International Conference on Information Systems (ICIS 2022)
Publication type	Conference Proceeding
Publication year	-
Rating (Jourqual 3)	A
Publication status	Submitted

Abstract

Although conversational agents (CA) are increasingly used for providing purchase recommendations, important design questions remain. The impact of recommendation modality, i.e., speech vs. text, and the optimal modality depend on how the recommendation outcome is explained. Across two experiments we examine with a novel fluency mechanism how modality shapes recommendation evaluation (persuasiveness and risk), the intention to follow the recommendation and how modality interacts with the style of recommendation explanation (verbal vs. numerical). Findings provide robust evidence that text-based outperform speech-based CAs in terms of processing fluency and consumer responses. They show that numerical explanations are a crucial design aspect to increase fluency and purchase intention of both CAs. We show the importance of processing fluency for the decision to follow a recommendation and highlight that fluency can be actively shaped through CA design. For practice, we offer actionable implications on how to make effective sales agents out of CAs.

Keywords

Recommendation Modality, Conversational Agent, Recommendation Explanation, Processing Fluency, Purchase Intention

F.1 Introduction

Firms increasingly use conversational agents (CA) based on artificial intelligence (AI) for providing purchase recommendations in order to influence consumers' decision making (Chernev, 2003) and boost sales (Karnik, 2019). For example, many retailers such as Sephora employ CAs to provide virtual product consultancy and shopping assistance (Arthur, 2016). CAs represent a novel interaction paradigm between consumers and firms and has been coined the “next operating system in commerce” (Tricks, 2021). They allow firms to engage consumers in an immediate, dialogue-based interaction to provide information on products and services (Miao et al., 2022). The provision of recommendations by CAs can occur either through speech or text (Wirtz et al., 2018), with the number of “speech commerce” users having grown by 120% in the last three years (Kinsella, 2020). Given these options, it is crucial for firms to understand how the modality of providing purchase recommendations (speech- vs. text-based) influences the chance that consumers follow the recommendation and purchase a product. However, research on how different recommendation modalities impact downstream consequences for consumers (information processing and perceptions) and for firms in turn (purchase) and how recommendations should be “designed” for different modalities has been scarce. So far, modality-related research in the technology context has focused on the influence of modality on preference expression (Klesse et al., 2015), on product search (King et al., 2021), as well as on the suitability of different modalities for different consumer tasks (Rzepka et al., 2021) and communication styles (Whang & Im, 2021).

We long know from communication theories (i.e., media richness, media synchronicity) that speech- and text-based modalities differ fundamentally in their characteristics (Schmitt, Zierau, et al., 2021). Specifically, speech is more synchronous and provides a richer symbol set than text (Moffett et al., 2021). It is therefore likely that the same message is processed differently when read than when heard. For instance, in the advertising context, it has been shown that communication modalities provide different metacognitive experiences, ranging from fluency to disfluency (Fransen et al., 2010). Thus, depending on the modality, certain purchase recommendations are more likely to be processed smoothly than others. In recommendation situations with CAs, where consumers are exposed to a relatively novel and uncertain decision environment, whether recommendation information are easily accessible can make a difference for the purchase decision. It is therefore essential for firms to understand how modality influences

the processing of purchase recommendations so that consumers feel well advised in their product choice and in turn purchase the recommended product.

However, for consumers' decision to follow the purchase recommendation, a crucial issue is that they understand why a CA provided the recommendations (Cheng & Jiang, 2022). A widely proposed mechanism for creating transparent purchase recommendations is to provide explanations for the recommendations (Wang & Benbasat, 2007). Explanations have been shown to support consumers to effectively use CAs (Wang & Benbasat, 2007) and to make better product decisions (Gregor & Benbasat, 2006). In the context of e-commerce, explanations can be communicated in a numerical or verbal style. Numerical explanations include information that contains specific numbers, e.g., "87% of consumers purchased this product" while verbal explanations consist of information that includes general descriptions, e.g., "many of the consumers purchased this product". As different explanation styles are processed and remembered differently (Viswanathan & Childers, 1996), it is likely that certain combinations of recommendation modality and explanation style are more easily processed than others. To date, no study has examined whether aligning recommendation modality and recommendation explanation style makes a difference for purchase intentions.

Taken together, this paper aims to address the following two research questions:

How does the modality for providing recommendations through conversational agents influence consumers' interaction perceptions and purchase intention?

How does the interplay of recommendation modality and explanation style influence consumers' interaction perceptions and purchase intention?

The structure of the paper is as follows: Initially, we explain processing fluency theory as the conceptual underpinning of this study. Then, we introduce the previous state of research on the two key constructs recommendation modality and explanation of recommendations. Subsequently, we present the research framework and develop the hypotheses. Our research framework was tested in two empirical studies. For this purpose, we outline study design, samples and measures of the two studies. The results are next presented and discussed. We complete the paper with theoretical and practical contributions, limitations, and future research directions.

F.2 Theoretical Background and Related Work

The following section provides the theoretical background for our research. Therefore, we derive the relevance of processing fluency for facilitating interaction evaluations and

purchase intention in the context of interactions with CA and elaborate how processing fluency is shaped by recommendation stimuli (modality and explanations). Further, we present related work on recommendation modalities and explanations of recommendations in the service context.

F.2.1 The Role of Processing Fluency in Purchase Contexts

Processing fluency describes the subjective feelings of ease or difficulty that consumers experience when mentally processing information (Graf et al., 2018; Novemsky et al., 2007). Processing fluency is a central construct that influences consumer evaluations and behavior across a wide array of domains (Schwarz et al., 2021). For example, studies have examined how aesthetic aspects such as readability of information or product design (Graf & Landwehr, 2015; Heitmann et al., 2020; Sung et al., 2022), consumer online ratings (Kostyk et al., 2017), product arrangements (Walter et al., 2020), and presentation of product information (Fang et al., 2007) affect processing fluency. A concept somewhat related to fluency is cognitive effort, i.e., the perceived time and effort required to complete a specific task (le Bigot et al., 2007; Mosteller et al., 2014; Rzepka et al., 2021). However, the consequences of cognitive effort for consumer responses depend on the level of fluency of information processing associated with these cognitive investments (Mosteller et al., 2014). Hence, high cognitive effort for a task need not be detrimental for consumer responses if it entails high fluency of information processing. Therefore, we argue that the level of processing fluency underlying a CA interaction might a better mechanism for explaining the effects of CAs on consumer responses. In line with this argument, extant research on how speech- vs. text-based CAs impact cognitive effort and consumers' subsequent responses provide inconclusive findings, likely because the different degrees of processing fluency underlying the interaction with the examined CAs and the resulting cognitive effort varied across these studies. No study so far has examined whether the modalities of interactions with CAs have differential influence on processing fluency (Schwarz et al., 2021). We address this research gap by examining the impact of the modality of CAs, specifically relating to the provision of recommendations, on processing fluency.

Further, we argue that two key recommendation evaluations strongly influenced by processing fluency are recommendation persuasiveness and the risk of adopting a recommendation. Only when consumers are persuaded by a recommendation they are willing to buy a product (Cialdini, 2009; Rhee & Choi, 2020). While consumers may be per-

suauded of the purchase recommendation's fit, they may still feel a certain risk of adopting the recommendation, either socially or financially, for instance. This in turn may prevent them from buying a recommended product (Tsiros & Heilman, 2005).

As already mentioned, the implementation and design of explanations are relevant in the recommendation context of different CAs. Previous fluency research that examined explanations has focused on online consumer ratings presented in a mean vs. distribution format (Kostyk et al., 2017). However, no study so far investigated whether giving an explanation has any effect at all on the processing fluency of recommendations and subsequent recommendation perceptions (persuasiveness and risk) and whether there is a difference if it is communicated in a numerical or verbal style.

F.2.2 Related Work on Recommendation Modalities

Until now, most studies on CAs have focused on the design and acceptance of text-based agents (Araujo, 2018; Hill et al., 2015; Huang & Rust, 2018), but recently, the emergence of speech-based interactions in commercial activities creates new challenges for the effective presentation of purchase recommendations. Firms are particularly interested in determining whether and how to implement speech-based recommendations (Bentahar, 2018). Specifically, the communication-related affordances of speech-based interactions differ fundamentally to text-based modalities that are likely to differentially affect consumer perceptions of the purchase recommendation process.

According to Moffett et al. (2021), modalities differ regarding the degree of synchronicity. Speech-based recommendation modalities are characterized by high degree of synchronicity as speech-based communication creates an immediate interaction scenario, in which consumers are involved in a constant back and forth. In a speech-based interaction, consumers must adapt to the speed of the other. In addition, consumers cannot revise and refine a message; speech-based interactions require an immediate, unfiltered response (Schmitt, Wambsganss, et al., 2021). In contrast, text-based interactions typically allow for a delayed communication style and thus have a low degree of synchronicity. In text communication, consumers can read and understand another's message at their own speed and compose their response in the same way (le Bigot et al., 2007). Furthermore, modalities differ in terms of the richness of symbols. Speech-based interactions provide a richer symbol set than text-based interactions. Specifically, speech-based interactions allow for the integration of vocal features (i.e., changes in frequency and amplitude) and prosody (i.e., pauses). To that end, according to media richness theory, speech-based interactions are “richer” than text-based interactions and

afford consumers to send a variety of social cues making communication more efficient (Rzepka et al., 2021).

Despite these fundamental differences between speech and text modality, there is limited research on how speech-based interactions determine the effectiveness of recommendations. The emerging literature on speech-based recommendation has focused on mechanisms related to anthropomorphism and its downstream consequences for consumers and firms, showing that speech-based recommendations can enhance social experiences in e-commerce settings (Hess et al., 2009; Qiu & Benbasat, 2009; Qiu & Benbasat, 2005). However, research on how the affordances of speech-based CAs affect information processing is scarce. Initial research has shown that speech-based preference expression can lead to more indulgent product choice when ordering a product with a vending machine (Klesse et al., 2015), hinting at modality-induced differences in consumer decision making processes. Looking at search modalities, King et al. (2021) show that speech-based search decrease purchase intention based on a more deliberative mindset. Similarly, taking a task-technology-fit perspective, Rzepka et al. (2021) could show that depending on the task type speech- vs. text-based CA differentially effect the level of effort (and enjoyment) associated with a service encounter when receiving a restaurant recommendation. Building on this work, we aim to investigate how the modality of a recommendation affects the cognitive processing of a purchase recommendation. Specifically, as our review of (recommendation) modalities and related work has indicated that speech-based compared to text-based interaction imposes considerable limitations in terms of quantity, structure, and processing time of information (Rzepka et al., 2021; Whang & Im, 2021), we focus on processing fluency as an important psychological mechanism, which has already been shown to be a potent theory to explain the effectiveness of message design across contexts.

F.2.3 Related Work on Explanations of Recommendations

CAs are able to provide customized recommendations based on the needs and desires of consumers articulated during an interaction. Firms increasingly provide recommendations in product-related CA interactions to use such agents not only for providing information and services but to sell products (Karnik, 2019). For instance, they often recommend an appropriate product to consumers from a wide range of choices and thus cognitively ease consumers' workload and facilitate purchase (Aksoy et al., 2006).

In the context of CAs, a recommendation is based either on collaborative filtering, i.e., products are recommended on the basis of similar consumers, or on content-based filtering, i.e., products are recommended on the basis of the consumer's preferred products (Gai & Klesse, 2019). Firms often use a hybrid form of this filtering for their algorithm. Current research therefore focuses on the impact of the recommendation algorithm (collaborative- vs. content-based vs. hybrid). However, consumers can only understand how the algorithm and CA work by explaining how a recommendation was made. Research has shown that explanation about the provided recommendation leads to higher trust in CAs (Wang & Benbasat, 2007). In e-commerce, collaborative explanations are typically used for recommendations in search tasks, since consumers do not have a reference value and thus find it more difficult to question the CA's statements (M. Liao & Sundar, 2021). Since the search task is currently the most common task in the recommendation context with CAs, our work also focuses on collaborative explanations. Therefore, two fundamentally distinct styles of communicating such explanations frequently used by firms are numerical and verbal communication. These are two different heuristics that consumer processes differently (Viswanathan & Childers, 1996). Numerical information like "83% of consumers also bought..." encompasses specific attributes while verbal information like "most of the consumers also bought..." comprises a generic description of attributes (Viswanathan & Childers, 1996). However, up to now, no research exists on which explanation style is most effective in terms of enhancing the fluency of processing the explanation. Moreover, it is unknown so far whether a certain explanation style fits a certain modality better than others. We contribute to this gap and examine which recommendation explanation style should be used for a specific modality.

F.3 Hypotheses Development

In this section, we first introduce the research framework based on the previous theory as well as literature review and develop the single hypotheses of the research model. The hypotheses development follows in three steps. Initially, we infer the effect of recommendation modality on processing fluency and in turn on recommendation evaluation and purchase intention. These hypotheses are tested in study 1. In a second step, we hypothesize the effect of explanation of recommendation on consumer responses, while in a third step, we elaborate the effects of the interaction between recommendation modality and explanation of recommendation. These effects regarding the second and third steps are tested in the second study.

F.3.1 Research Framework

In order to understand what impact our three stimuli (recommendation modality, recommendation explanation and their interplay) has on processing fluency of the recommendation and what influence this central mechanism has on the subsequent evaluation of the recommendation and finally on the behavioral intention, we established the following research model, which is presented in Figure 21. To analyze the influence of the recommendation modality and recommendation explanation as well as their interaction, they were added as independent variables to the research model. The processing fluency in turn mediates the effect of the three stimuli on the recommendation persuasiveness and on the risk of adopting the recommendation, which in turn influence behavioral response in terms of purchase intention. Hence, consumers' psychological processes capture the direct processing of the stimuli in a fluent or disfluent manner and the resulting evaluation of perceived persuasiveness and risk of the provided recommendation.

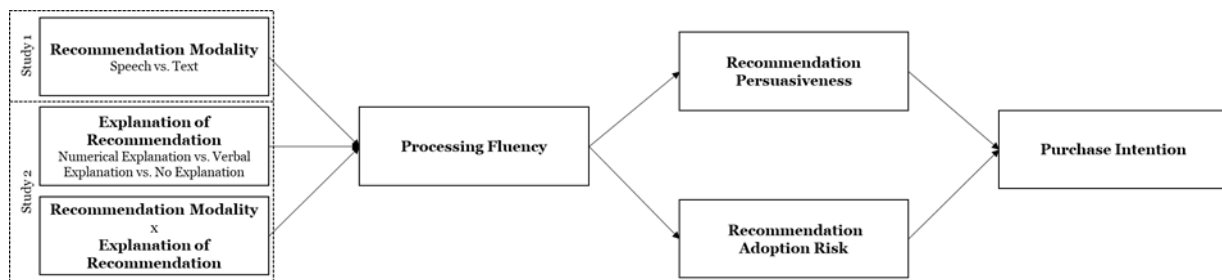


Figure 21. Research model

F.3.2 The Effect of Recommendation Modality of Perceptions and Purchase Intention

CAs who provide purchase recommendations typically reduce the amount of information by minimizing the selection of recommended products as well as sharing only the most important details of the products. Condensing information increases complexity and mental workload within speech-based interaction as it is more difficult to listen than to read (Hong et al., 2004; le Bigot et al., 2007). Different presentation styles of information (simultaneous vs. sequential) can also increase or decrease complexity (Basu & Savani, 2019; Mogilner et al., 2013). Simultaneous presentation of information can lower the difficulty of comparison (Basu & Savani, 2019). Simultaneous presentation is a feature of text-based interaction that eases information processing (Schmitt, Wambsganss, et al., 2021). In particular, consumers find it more difficult to assimilate and remember information via speech (le Bigot et al., 2007), especially when comparing

between options. When CAs give purchase recommendations, a comparison of at least two product options occurs. For this type of task, the modality of text seems to fit better as it presents information simultaneously and allows the consumer to compare the options. Thus, we obtain the following hypothesis:

H1: Consumers perceive the processing of recommendations by a speech-based conversational agent (compared to a text-based conversational agent) as less fluent.

Processing fluency is described as a subjective feeling that leads consumers to make judgments and show behaviors in response to this good or bad feeling of fluency or disfluency (Schwarz et al., 2021). Research findings indicate that stimuli that are perceived as easy to process are also perceived as more positive by consumers (Winkielman et al., 2003). The positive feeling leads to a more reliable evaluation of information and a certain ease within the decision process which increases the general liking of a certain information (Mosteller et al., 2014; Schwarz et al., 2021), which conceivably leads consumers to be persuaded by the obtained recommendation. Hence, a positive feeling can be seen as a heuristic that consumers use to evaluate the persuasiveness of a situation (Chen et al., 1999). To ensure that consumers consider a recommendation as persuasive, recommendations need to convey arguments or cues that influence consumer evaluations and responses (Cialdini, 2009; Dehnert & Mongeau, 2022). For instance, the positive feeling of processing fluency gives the impression that a recommendation is beneficial for the consumer (Kahneman & Frederick, 2007). In other words, it is convincing and leads to a certain behavior, in the case of purchase recommendations, to buy a recommended product. This leads us to the following two hypotheses:

H2a: If processing fluency is high, persuasiveness of recommendations increases.

H2b: If persuasiveness of recommendations is high, purchase intention increases.

Furthermore, the ease of processing information can contribute to a feeling of familiarity and thus reduce the perceived risk of a hazardous situation or decision (Schwarz et al., 2021). Thus, the ease of processing information can lead to a perceived sense of security (Dohle & Montoya, 2017). Also, consumers do not want to lose the positively acquired feeling of ease, so that they assess a situation as less risky, thus avoiding losses. Consumers therefore judge the acceptance of recommendations as less risky and are thus more likely to purchase a product in order not to lose the positive feeling of fluency (Kahneman & Frederick, 2007). We therefore hypothesize the following:

H3a: If processing fluency is high, risk of recommendation adoption decreases.

H3b: If risk of recommendation adoption is high, purchase intention decreases.

F.3.3 The Effect of Explanation of Recommendation on Processing Fluency

The communication of an explanation for recommendations can be considered as a cue to help consumers to make decisions. This type of stimuli signals to consumers that they receive a recommendation that is based on the analysis of their needs. As shown in the theoretical framework, there are different styles of explanations that can be provided to a consumer. As already mentioned, firms use the collaborative approach for explanations (i.e., explanations referring to what other consumers did) in particular, due to the benefits in the search context of products (M. Liao & Sundar, 2021). Moreover, collaborative explanations fuel the bandwagon heuristic. As Sundar (2008) outlines, consumers like to go with the flow of people and assume that if a large number of their fellows agree that something is good, then consumers will be more likely to adopt that mindset. This kind of heuristic leads consumers to be able to make decisions without thinking much about it (Kahneman & Frederick, 2007). As a consequence, giving any explanation for the recommendation conveys a kind of reliability to the consumer (M. Liao & Sundar, 2021), leading to the result that consumers do not have to question the recommendations, so the recommendations will be processed with an increased ease. Thus, we obtain the following hypothesis:

H4: A (a) numerical and (b) verbal explanation of the recommendations (vs. no explanation) increases processing fluency.

Due to the effects of processing fluency on perceived recommendation persuasiveness, recommendation adoption risk and purchase intention explained in the previous section, and the assumption that explanations have a positive effect on processing fluency, explanations of the recommendations also have a positive effect on the subsequent perception and outcome variables.

F.3.4 The Interplay of Recommendation Modality and Explanations

As mentioned above, in collaborative explanations for a purchase recommendation, firms relate to a reference group (i.e., other consumers) which is either expressed with specific numbers like "XX% of consumers similar to you" or descriptive words like "most consumers similar to you" (M. Liao & Sundar, 2021).

Because of their specificity, consumers can better distinguish numerical information from other information and process it more easily than verbal information. However, numerical information often do not have inherent meaning, whereas verbal information can more effortlessly convey meaning (Viswanathan & Childers, 1996). Therefore, numerical information require greater cognitive effort, making these information easier to

be overlooked, particularly comparing this to vivid or verbal information (Zillermann & Brosius, 2012), which can convey meaning more easily. However, presenting numerical information in text-based form can perform better than verbal information for product comparisons (Viswanathan & Childers, 1996). The higher degree of accurate encoding in comparison tasks offsets the disadvantage of transforming numerical information to extract its inherent meaning. In text-based interactions, more complex information can be processed than in speech-based interactions, while the latter is better suited for conveying simpler (i.e., verbal) information (Schmitt, Wambsganss, et al., 2021). Accordingly, we assume that a numerical explanation is considered easier to process in text-based interaction and a verbal explanation is easier to process in speech-based interactions, thus the following hypotheses emerge:

H5a: A numerical explanation (vs. verbal explanation) of the recommendations in a text-based interaction with a conversational agent increases processing fluency.

H5b: A verbal explanation (vs. numerical explanation) of the recommendations in a speech-based interaction with a conversational agent increases processing fluency.

F.4 Experimental Procedures and Research Context

To test our research model, we used a custom-made conversational interface building on *Python* and *Google WaveNet* (Zierau et al., 2022). The interface provides the ability to model both speech- and text-based conversational interactions. Across modalities, the design of the interface was kept simple to control for potential confounds and to isolate the effect of recommendation modality on consumer perceptions (see Figure 22). for a comparative overview of both interfaces). For both studies, the dialogue flow is predefined to ensure that the content and structure of the conversational interaction is kept stable across participants. For the speech-based interface, we used a state-of-the-art text-to-speech generator from Google WaveNet to create the different voice prompts (Oord et al., 2016).

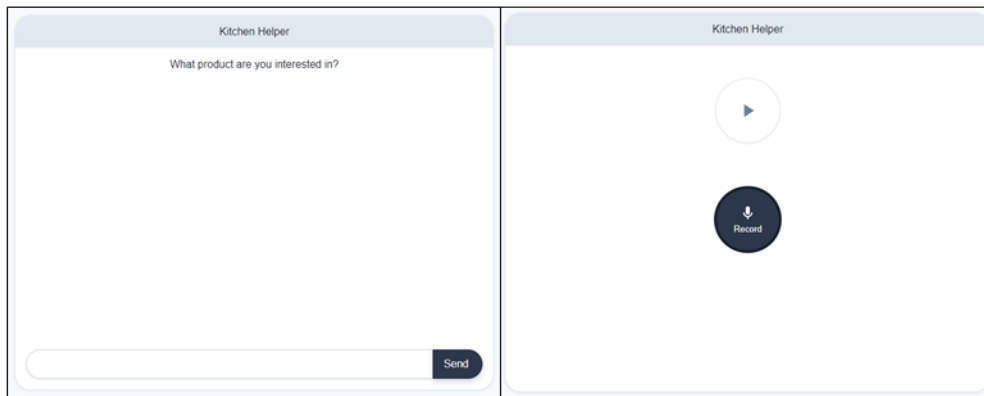


Figure 22. Text-based vs. speech-based CA

As a study context, we chose purchase recommendations within the food industry for two major reasons. First, one of the first contexts in which purchase recommendations have been made via a speech-based CA is the food sector. Especially when cooking, the advantages of a speech-based CA can come into play in terms of hands-free operation (de Bellis & Venkataramani Johar, 2020; Sciuto et al., 2018). Second, consumers can imagine many food products and the level of complexity is low. Above all, it is not essential that the consumer needs a picture to evaluate the usefulness of a product. In recent years, numerous skills for speech- and text-based CAs have been developed to provide consumers with recipe tips and recommendations for food products. Thus, the scenario exhibits a high degree of realism, which enhances the external validity of our studies.

In this paper, we present evidence from two studies that were designed to test our hypotheses.

F.5 Study 1

Study 1 was designed to provide a first test of our baseline hypotheses that speech-based (as opposed to text-based) interactions evoke less processing fluency when getting a purchase recommendation. Furthermore, study 1 also tests whether increases in processing fluency increase recommendation persuasiveness and decrease the risk of adopting recommendations, and whether these changes increase purchase intention.

F.5.1 Design and Procedures

A total of 181 participants were recruited from a native English-speaking user panel (Prolific). We followed the guidelines of Lowry et al. (2016) for the recruitment of the participants. Those who failed to answer attention or treatment checks correctly were disregarded from further analyses. Thus, the effective sample resulted in a total of 141

participants ($M_{\text{Age}}=34.59$, $SD_{\text{Age}}=13.26$, 48.23% female). Participants were randomly assigned to a two-cell between subject experiment (speech- vs. text-based interaction). In both conditions, the experimental task comprised getting a purchase recommendation on peanut butter. Before the interaction with the conversational agent starts, we asked participants to immerse themselves in a scenario describing an event, where they were preparing a desert for their friends. In the scenario, we provided them with a set of preferences related to product features (e.g., maximum price), they were instructed to report to a fictitious app called *KitchenHelper* (via speech or text depending on the condition). We asked participants to report all of these preferences to receive a purchase recommendation. The sequence and semantics of the questions the interface was posing to participants were identical across conditions and participants (i.e., the sequence of eight conversational turns, semantics, and all other features of the purchase recommendation process). The participants were only allowed to use laptops or desktop computers to avoid confounding affects based on device.

Immediately after the experimental task, we assessed participants perception of processing fluency using a 4-item scale (adapted from Graf et al., 2018); sample item: “*The process of studying food product recommendations of the agent was: disfluent / fluent*”: 7-point differential Likert scale; $\alpha_{\text{Fluency}} = .91$). Moreover, we assessed recommendation persuasiveness using 4 items (adapted from (Shevchuk et al., 2019); sample item: “*The food product recommendations of the agent had an influence on me.*”: 7-point Likert scale, from 1: “Strongly disagree” to 7: “Strongly agree”; $\alpha_{\text{Persuasiveness}} = .81$) and recommendation adoption risk using a 4-item scale (adapted from H. Song & Schwarz, 2009; sample item: “*How did you rate the risk of adopting a food product recommendation from the agent?: very safe / very risky*”: 7-point differential Likert scale; $\alpha_{\text{Risk}} = .93$). Finally, we measured purchase intention using 3 items (adapted from Bleier et al., 2019 and King et al., 2021; sample item: “*It is likely that I would purchase at least one of the recommended food products I searched for.*”: 7-point Likert scale, from 1: “Strongly disagree” to 7: “Strongly agree”; $\alpha_{\text{PurchaseIntention}} = .87$). We captured a battery of control measures including measures on social-demographics, scenario realism, and personal traits and experiences with CAs.

F.5.2 Results

To test H_1 , we conducted a one-way analysis of variance (ANOVA) revealing that participants who used the speech-based CA experienced significantly lower levels of processing fluency compared to participants using the text-based CA ($M_{\text{Text}} = 6.43$, $SE = 0.13$; $M_{\text{Speech}} = 5.93$, $SE = 0.12$; $t = -7.64$, $p < 0.01$). Thus, we can confirm H_1 .

Furthermore, we conducted a seemingly unrelated regression (SUR). Since the dependent variables of the processing fluency model are simultaneously the independent variables of the purchase intention model, it seems unrealistic for them to have independent error terms because they are related in content. Therefore, the method of SUR is able to provide more efficient estimates for coefficients (Zellner, 1962). The path model results are summarized in Figure 4. These findings demonstrate that the decrease of processing fluency evoked by the speech-based CA ($\beta_{\text{Speech}} = -.50$, $z = -2.83$, $p < .01$) in turn led to a significant decrease in consumers' evaluation of recommendation persuasiveness ($\beta_{\text{Fluency}} = .52$, $z = 5.62$, $p < .001$) and a significant increase in consumers' evaluation of recommendation adoption risk ($\beta_{\text{Fluency}} = -.50$, $z = -5.19$, $p < .001$). Focusing on the impact on purchasing intention, we found that only the negative impact of speech-based CA on recommendation persuasiveness, in turn, led to significant decrease in purchase intention ($\beta_{\text{Persuasiveness}} = .87$, $z = 9.14$, $p < .001$), while the increase in recommendation adoption risk had no significant impact on purchase intention ($\beta_{\text{Risk}} = .02$, $z = .19$, $p = .86$). Thus, we were able to confirm hypotheses H_{2a} , H_{2b} , and H_{3a} and had to reject hypothesis H_{3b} .

F.5.3 Discussion

The findings of Study 1 provide initial evidence for our baseline hypothesis that speech-based CA create less processing fluency compared to text-based CA. Moreover, we can show that these differences lead to reduced recommendation persuasiveness and increased risk evaluation when using a speech-based CA. However, only recommendation persuasiveness can be linked to purchasing intention.

F.6 Study 2

The main objective of Study 2 was to examine to which extent the presence of an explanation for the recommendation can either enhance or reduce participants' perception of processing fluency. Moreover, we distinguished between explanations that either used a numerical or a verbal style.

F.6.1 Design and Procedures

Again, a total of 362 participants were recruited from a native English-speaking user panel (Prolific), following the guidelines of Lowry et al. (2016). We disregarded those participants that either failed to answer our attention checks correctly and those who did not pass the treatment checks. Specifically, those who did not recognize whether the explanation used a numerical vs. verbal style were disregarded from the sample. The final sample consisted of 293 participants ($M_{\text{Age}}=32.62$, $SD_{\text{Age}}=12.63$, 56.66% female). Participants were randomly assigned to a 2 (recommendation modality: speech-based vs. text-based) x 3 (explanation: numerical explanation vs. verbal explanation vs. no explanation) between-subjects design, using the same experimental paradigm and context as in Study 1 and only adding a short explanation. To distinguish between the different explanation conditions, we kept the content and meaning stable and only shifted between a numerical and verbal communication style. Specifically, the numerical condition used numbers (i.e., “*Based on preferences of 91.8% of the Kitchen Helper App users and about 652,000 reviews of around 17 peanut butter products, [...]*”), the verbal condition used words (i.e., “*Based on preferences from most of the Kitchen Helper App users and a lot of reviews of many peanut butter products, [...]*”). For the no explanation condition, we added a short sentence stating that the CA had considered all available options, but did not offer any explanation. To control for the length of the turn and to isolate the effect of the explanation, we held the number of words constant between conditions. Immediately, after the experimental task we collected the same measures as in Study 1 for processing fluency ($\alpha_{\text{Fluency}} = .91$), recommendation persuasiveness ($\alpha_{\text{Persuasiveness}} = .82$), recommendation adoption risk ($\alpha_{\text{Risk}} = .93$), and purchase intention ($\alpha_{\text{PurchaseIntention}} = .85$).

F.6.2 Results

To test H₁, H₄ and H₅, we conducted a one-way analysis of variance (ANOVA). The results of ANOVA show a significant main effect of recommendation modality on processing fluency ($F(1, 287) = 8.31$, $p < 0.01$). Planned contrasts of predictive margins show that speech is significantly less fluent than text ($M_{\text{Speech}} = 5.74$, $SE = 0.09$; $M_{\text{Text}} = 6.12$, $SE = 0.09$; $t = -2.88$, $p < 0.01$), which provides support for H₁. Also, the main effect of providing an explanation on processing fluency was significant ($F(2, 287) = 8.90$, $p < 0.001$). Again, planned contrasts of predictive margins show that numerical explanation is significantly more fluent than no explanation ($M_{\text{NoExplanation}} = 5.86$, $SE = 0.12$; $M_{\text{NumericalExplanation}} = 6.29$, $SE = 0.11$; $t = 2.73$, $p < 0.01$). But there is no significant

difference between verbal explanation and no explanation ($M_{\text{VerbalExplanation}} = 5.64$, $SE = 0.12$; $t = -1.35$, $p = 0.180$). Therefore, we could confirm H_{4a} and had to reject H_{4b} . We use planned contrasts to analyze the interaction effect of recommendation modality and explanation of recommendation. Figure 23 shows the predictive margins of the interaction effects. We find support for H_{5a} , as the processing fluency is higher for numerical explanations than for verbal explanations provided by text-based CAs ($M_{\text{numericalXtext}} = 6.54$, $SE = 0.16$; $M_{\text{verbalXtext}} = 5.72$, $SE = 0.16$; $t = 3.59$, $p < 0.001$). However, we find no support for H_{5b} . Instead, we identified a reversed effect, as the numerical explanation has a significantly higher effect on fluency than the verbal explanations provided by speech-based CAs ($M_{\text{numericalXspeech}} = 6.05$, $SE = 0.14$; $M_{\text{verbalXspeech}} = 5.56$, $SE = 0.17$; $t = 2.21$, $p < 0.05$).

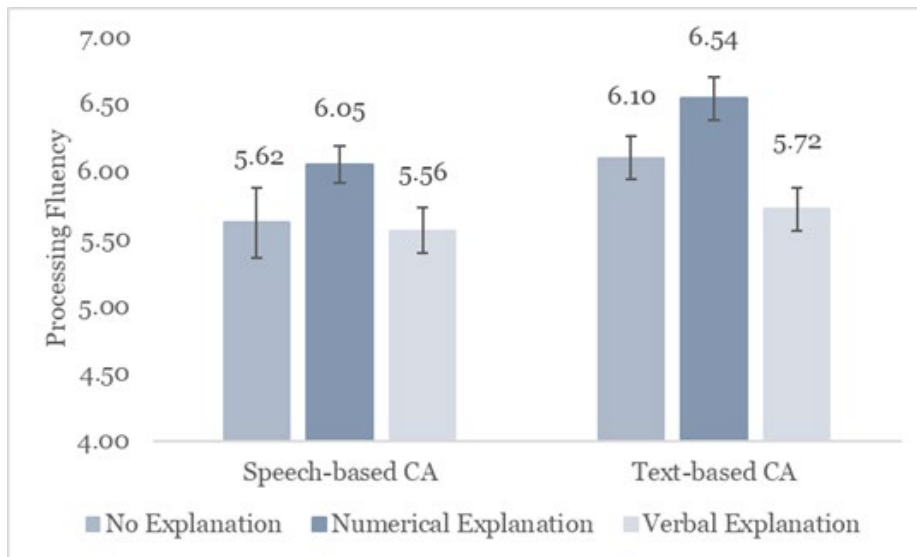
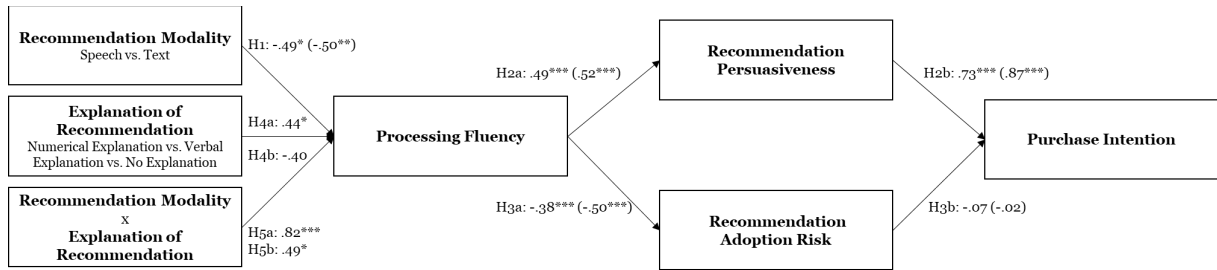


Figure 23. Predictive margins of the interaction effects for processing fluency

Next, we estimated a path model using the SUR function in STATA to assess the overall system of hypotheses, see Figure 24. These findings demonstrate that the decrease of processing fluency evoked by the speech-based CA ($\beta_{\text{Speech}} = -.49$, $z = -2.14$, $p < .05$) in turn led to a significant decrease in consumers' evaluation of recommendation persuasiveness ($\beta_{\text{Fluency}} = .49$, $z = -8.31$, $p < .001$) and a significant increase in consumers' evaluation of recommendation adoption risk ($\beta_{\text{Fluency}} = -.38$, $z = 4.92$, $p < .001$). Focusing on the impact on purchasing intention, we found that only the negative impact of speech-based CA on recommendation persuasiveness, in turn, led to significant decrease in purchase intention ($\beta_{\text{Persuasiveness}} = .73$, $z = 12.10$, $p < .001$), while

the increase in recommendation risk had no significant impact on purchase intention ($\beta_{\text{Risk}} = -.07, z = -1.45, p = .15$). The results of the SUR also show that a numerical explanation (vs. no explanation; $\beta_{\text{NumericalExplanation}} = .44, z = 1.99, p < .05$) increases the processing fluency significantly, but a verbal explanation (vs. no explanation) had no effect ($\beta_{\text{VerbalExplanation}} = -.40, z = -1.78, p = .1$). As in study 1, we were able to confirm hypotheses H_{2a}, H_{2b}, and H_{3a} and had to reject hypothesis H_{3b}.



Notes: $*p \leq 0.05$; $**p \leq 0.01$; $***p \leq 0.001$; text and values in parentheses refer to study 1

Figure 24. Path model of Study 1 & 2

F.6.3 Discussion

The results of study 2 provide further evidence for our initial hypothesis that speech-based interactions produce lower processing fluency compared to text-based interactions. In addition, we repeatedly demonstrate that processing fluency reduces the risk evaluation of recommendation adoption and increases recommendation persuasion. Recommendation persuasion also increases purchase intention, whereas risk evaluation does not, as in Study 1. Thus, processing fluency is an important mechanism that determines whether consumers purchase a recommended product. Moreover, numerical explanation of recommendations leads to higher processing fluency of the recommendation for both speech- and text-based CAs.

F.7 General Discussion

Firms face the challenge to decide which type of CAs they want to use for communicating purchase recommendations, because it is well established that CAs can help consumers to make better product decisions (Gai & Klesse, 2019). Our aim was to show how fluently purchase recommendations can be processed via speech- and text-based CAs and what effects this has on the evaluation of the recommendations and the corresponding purchase intentions. In addition, we aimed to show if an explanation of the recommendations can contribute to the ease of processing purchase recommendations.

Since speech- vs. text-based interactions comprise fundamental differences (Schmitt, Zierau, et al., 2021), and yet explanations are a central component of recommendations, it is even more relevant to understand whether and how explanations have an effect on processing fluency within different recommendation modalities.

We showed in both studies that purchase recommendations provided by text-based CAs are perceived as easier to process as compared to speech-based CAs. Speech-based interactions are frequently hyped as a new technology due to their naturalness and flexibility, leading to a growing interest in integrating them in the purchasing context (Schwartz, 2022). However, shopping, especially for new products, is more complex in nature, as consumers want to feel they are being well advised by receiving customized recommendations as well as understanding why they are receiving certain products from a CA (Gai & Klesse, 2019; Wang & Benbasat, 2007). Across two studies, our results demonstrate that even when only two purchase recommendations need to be compared, the text-based CA outperforms the speech-based CA.

The processing fluency of purchase recommendations is a central mechanism, which primarily conveys the feeling of whether recommendations can be processed as easy or difficult. This feeling in turn leads to an evaluation of the recommendation and the intention to follow the purchase recommendation by purchasing the product(s). Our results show that processing fluency, which is associated with feelings of ease, significantly minimized perceived risk and increased persuasiveness. We also found that perceived persuasiveness significantly increased the intention to purchase a recommended product. However, risk of adopting a recommendation did not significantly influence the purchase intention. A possible explanation for the non-significant effect is the food product used in our experiment. Food products such as peanut butter are low-involvement products for which the risk of purchasing the recommended product is manageable. However, in the context of shopping with CAs, low-involvement products are currently the standard.

The second study examined a further stimulus on processing fluency, the explanation for the recommendations. Furthermore, we examined whether there was an interaction effect between the type of CA and the explanation for recommendations. A numerical explanation was significantly perceived as more fluent than no explanation. However, a verbal explanation was not perceived as more fluent than no explanation, i.e., there were no significant differences between the two styles. Thus, we note that a numerical explanation is more specific than a verbal explanation, which is more generically in nature (Viswanathan & Childers, 1996). Not necessarily processing ability, but processing

preferences may make a difference in whether and which style of explanation is preferred by the consumer (Childers et al., 1985). Since consumers are confronted with numbers on a regular basis when making product decisions and since these numbers are more unambiguous and easier to interpret, a numerical explanation is suitable for processing purchase recommendations more easily. Interestingly, numerical explanations (as opposed to verbal explanations) have a significantly higher influence on processing fluency for both speech- and text-based CAs. In particular, numbers can be identified as a heuristic and a kind of anchor. Processing numbers in speech-based interactions requires some cognitive effort, yet these numbers do not need to be internalized in their entirety (Viswanathan & Childers, 1996), and still convey a sense of accuracy when the numbers are highly marked (M. Liao & Sundar, 2021; Schmitt, Wambsganss, et al., 2021). The feeling of seeing or hearing a high number conveys the feeling of an easier processing of the purchase recommendations, because the recommendations can be better categorized.

Our two studies have shown a negative effect of speech-based CAs compared to text-based CAs in e-commerce. Firms should therefore carefully consider to prioritize a text-based CA for digital commerce over a speech-based CA or, alternatively, consider how they can increase the fluency of the speech-based CA. Previous studies, primarily focused on social presence and anthropomorphic aspects of CAs, have highlighted the benefits of speech-based CAs and their social meaning, which could also be used in context of purchase recommendations. Therefore, further design-oriented questions arise on how to raise these potentials in practice. The use of a numerical explanation can already increase the fluency of speech-based CAs, but also of text-based CAs, so that they are preferable in everyday commerce contexts in order to ultimately increase the purchase intention. Given the increase of a large number of available CAs in the omnichannel, such a generalizable conversational design is important, to provide superior consumer experiences across modalities.

F.8 Implications, Limitations and Future Research

Our study has several important implications for IS theory and practice. First, we provide a rigorous experimental approach to disentangle the effects of CA modality as one key material property of these IT artifacts. By highlighting across two studies that in shopping and purchase recommendation contexts, text-based CAs usually outperform speech-based CAs. Consequently, we challenge the assumption that naturalness of speech is superior compared to text-based interactions. Second, and well connected to

our first theory implication, we contribute with our fluency theory perspective to the better understanding of CA interaction. By highlighting the crucial role of fluent interaction on consumer perceptions, we underscore the impact that fluency theory could have in understanding on consumer reactions to CA interactions. Third, we highlight that not only fluency and modality matter for CA interaction outcomes. Research should also very carefully consider the content-wise perspective when highlighting crucial aspects for interactions, i.e., in our case recommendation explanations. By showing that numerical explanations outperform no explanations and verbal explanations, we identify important anchoring points during conversational interactions.

Consequently, we highlight as a major practical implication that firms should use in shopping contexts text-based CAs to communicate purchase recommendations and provide them a numerical explanation for the recommendations, so that the overall processing fluency is high, leading to the persuasion that the recommended product is appropriate and the consumer is more likely to purchase the product.

As advances in technology enable firms to implement both speech- and text-based CAs, and consumers' requests for these technologies increase, firms should consider how they will design each modality. Therefore, it is also important to take a multi- and omnichannel perspective. In the long term, it will not be a question of whether firms use speech- or text-based CAs, but rather consumers will decide which CA they prefer and when, so the design of both CAs will become equally relevant. However, our research has highlighted potential problems with speech-based CAs. Future research should therefore consider how these CAs can be designed more effectively so that the processing fluency of the recommendation increases.

Our paper has naturally further limitations that provide ground for future research. We conducted our two studies online by utilizing an online panel. Thus, we acknowledge that we could not control for all external stimuli in comparison to in-person lab experiments. Such noise could especially influence our central fluency variable. However, our setting is comparable to typical shopping experiences and, thus, we took this limitation consciously into account. Field experiments could thus take our findings as a starting point to provide evidence if our theory holds in large-scale shopping settings.

In the experiment, participants were asked to put themselves in a service encounter with the CA that was as realistic as possible. More specifically, this meant that they had to put themselves in a specific role (vignette), act accordingly, and make decisions. This type of experiment has limitations in external validity (Aguinis & Bradley, 2014).

Therefore, the study's results have limited generalizability to other or real-life situations in service encounters with CAs. Moreover, the experiment depicts a short-term and fixed process. Future experiments should also investigate more flexible service situations (along the lines of Wizard-of-Oz experiments) and their impact. Furthermore, future studies should test different situations in the context of purchase recommendation to strengthen the robustness of our results and discuss possible further effects.

When explaining a recommendation, we chose a collaborative explanation type in the experiment, however, there is also the variation of a content-based explanation, which should be further explored in future studies. Moreover, we designed the collaborative explanation using numerical and verbal elements. However, future studies should use alternative numerical and verbal elements to validate the results of our study. Furthermore, an explanation is a key design element in the recommendation context, but there are other design elements that should also be investigated. The sequence of recommendations, the presented product features, or social communication elements, etc. also comprise a recommendation and should be investigated in the context of speech- vs. text-based CAs.

References

- Abdolrahmani, A., Kuber, R., & Branham, S. M. (2018). Siri talks at you: An empirical investigation of voice-activated personal assistant (VAPA) usage by individuals who are blind. *ASSETS 2018 - Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*, 249–258. <https://doi.org/10.1145/3234695.3236344>
- Adamopoulou, E., & Moussiades, L. (2020). Chatbots: History, technology, and applications. *Machine Learning with Applications*, 2(November). <https://doi.org/10.1016/j.mlwa.2020.100006>
- Agarwal, R., & Karahanna, E. (2000). Time Flies When You're Having Fun: Cognitive Absorption and Beliefs about Information Technology Usage. *MIS Quarterly*, 24(4), 665–694.
- Aguinis, H., & Bradley, K. J. (2014). Best Practice Recommendations for Designing and Implementing Experimental Vignette Methodology Studies. *Organizational Research Methods*, 17(4), 351–371. <https://doi.org/10.1177/1094428114547952>
- Akahori, W., Sugiyama, H., Minami, H., Miyake, A., & Watanabe, M. (2019). Paired conversational agents for easy-to-understand instruction. *Conference on Human Factors in Computing Systems - Proceedings*, 1–6. <https://doi.org/10.1145/3290607.3312794>
- Aksoy, L., Bloom, P. N., Lurie, N. H., & Cooil, B. (2006). Should recommendation agents think like people? *Journal of Service Research*, 8(4), 297–315. <https://doi.org/10.1177/1094670506286326>
- Amazon. (2015). *Amazon echo now available to all consumers*. <https://press.aboutamazon.com/news-releases/news-release-details/amazon-echo-now-available-all-customers/>
- Antaki, C. (2008). Discourse analysis and conversation analysis. In *The SAGE Handbook of Social Research* (pp. 431–446).
- Araujo, T. (2018). Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions. *Computers in Human Behavior*, 85, 183–189. <https://doi.org/10.1016/j.chb.2018.03.051>

- Arthur, R. (2016). *Sephora Launches Chatbot On Messaging App Kik*. Forbes. <https://www.forbes.com/sites/rachelarthur/2016/03/30/sephora-launches-chatbot-on-messaging-app-kik/?sh=72577169547e>
- Ashktorab, Z., Jain, M., Vera Liao, Q., & Weisz, J. D. (2019, May 2). Resilient chatbots: Repair strategy preferences for conversational breakdowns. *Conference on Human Factors in Computing Systems - Proceedings*. <https://doi.org/10.1145/3290605.3300484>
- Bagozzi, R. P., Yi, Y., & Phillips, L. W. (1991). Assessing Construct Validity in Organizational Research. *Administrative Science Quarterly*, 36(3), 421–458. <https://doi.org/10.2307/2393203>
- Baird, A., Jørgensen, S. H., Parada-Cabaleiro, E., Cummins, N., Hantke, S., & Schuller, B. (2018). The perception of vocal traits in synthesized voices: Age, gender, and human likeness. *Journal of the Audio Engineering Society*, 66(4), 277–285. <https://doi.org/10.17743/jaes.2018.0023>
- Basu, S., & Savani, K. (2019). Choosing Among Options Presented Sequentially or Simultaneously. *Current Directions in Psychological Science*, 28(1), 97–101. <https://doi.org/10.1177/0963721418806646>
- Bem, S. L. (1981). *Bem Sex-Role Inventory: professional manual*. Consulting Psychologists Press.
- Benbasat, I., & Zmud, R. W. (2003). The Identity Crisis within the Is Discipline: Defining and Communicating the Discipline's Core Properties. *MIS Quarterly*, 27(2), 183–194.
- Benlian, A., Klumpe, J., & Hinz, O. (2020). Mitigating the intrusive effects of smart home assistants by using anthropomorphic design features: A multimethod investigation. *Information Systems Journal*, 30, 1010–1042. <https://doi.org/10.1111/isj.12243>
- Bentahar. (2018). *How voice shopping is changing shopping*. Forbes. <https://www.forbes.com/sites/forbesagencycouncil/2018/12/03/how-voice-search-is-changing-shopping/?sh=177940026152>
- Berger, A., Schlager, T., Sprott, D. E., & Herrmann, A. (2018). Gamified interactions: whether, when, and how games facilitate self–brand connections. *Journal of the Academy of Marketing Science*, 46(4), 652–673. <https://doi.org/10.1007/s11747-017-0530-0>

- Berger, J. A., Rocklage, M., & Packard, G. M. (2021). Expression Modalities: How Speaking Versus Writing Shapes What Consumers Say, and Its Impact. *Journal of Consumer Research*, *Forthcoming*. <https://doi.org/10.2139/ssrn.3791506>
- Berry, D. C., Butler, L. T., & de Rosis, F. (2005). Evaluating a realistic agent in an advice-giving task. *International Journal of Human Computer Studies*, *63*(3), 304–327. <https://doi.org/10.1016/j.ijhcs.2005.03.006>
- Bettencourt, L. A. (1997). Customer voluntary performance: Customers as partners in service delivery. *Journal of Retailing*, *73*(3), 383–406. [https://doi.org/10.1016/S0022-4359\(97\)90024-5](https://doi.org/10.1016/S0022-4359(97)90024-5)
- Bickmore, T., & Schulman, D. (2007). Practical Approaches to Comforting. *Proceedings of ACM CHI 2007: Conference on Human Factors in Computing Systems*, 2291–2296.
- Bickmore, T. W., Caruso, L., Clough-Gorr, K., & Heeren, T. (2005). “It’s just like you talk to a friend” relational agents for older adults. *Interacting with Computers*, *17*(6), 711–735. <https://doi.org/10.1016/j.intcom.2005.09.002>
- Bickmore, T. W., & Picard, R. W. (2005). Establishing and Maintaining Long-Term Human-Computer Relationships. *ACM Transactions on Computer-Human Interaction*, *12*(2), 293–327.
- Blake, M. (2018). *4 Ways to Disrupt Insurance with Customer Technology*. Forbes. <https://www.forbes.com/sites/blakemorgan/2018/05/02/4-ways-to-disrupt-insurance-with-customer-technology/?sh=467684b1382e>
- Bleier, A., Harmeling, C. M., & Palmatier, R. W. (2019). Creating effective online customer experiences. *Journal of Marketing*, *83*(2), 98–119. <https://doi.org/10.1177/0022242918809930>
- Bolt, R. A. (1980). “Put-That-There” : Voice and Gesture At the Graphics Interface. *Comput Graphics (ACM)*, *14*(3), 262–270. <https://doi.org/10.1145/965105.807503>
- Brandtzaeg, P. B., & Følstad, A. (2018). Chatbots: changing user needs and motivations. *Interactions*, *25*(5), 38–43.
- Brannon Barhorst, J., McLean, G., Shah, E., & Mack, R. (2021). Blending the real world and the virtual world: Exploring the role of flow in augmented reality experiences. *Journal of Business Research*, *122*, 423–436. <https://doi.org/10.1016/j.jbusres.2020.08.041>

- Braun, M., Mainz, A., Chadowitz, R., Pfleging, B., & Alt, F. (2019). At your service: Designing voice assistant personalities to improve automotive user interfaces a real world driving study. *Conference on Human Factors in Computing Systems - Proceedings*, 1–11. <https://doi.org/10.1145/3290605.3300270>
- Breivik, E., & Thorbjørnsen, H. (2008). Consumer brand relationships: An investigation of two alternative models. *Journal of the Academy of Marketing Science*, 36(4), 443–472. <https://doi.org/10.1007/s11747-008-0115-z>
- Burgoon, J. K., Manusov, V., & Guerrero, L. K. (2013). *Nonverbal signals*. Cengage Learning.
- Cafaro, A., Vilhjálmsón, H. H., Bickmore, T., Heylen, D., & Schulman, D. (2013). First impressions in user-agent encounters: The impact of an agent’s nonverbal behavior on users’ relational decisions. *12th International Conference on Autonomous Agents and Multiagent Systems 2013, AAMAS 2013*, 2, 1201–1202.
- Cambre, J., Reig, S., Kravitz, Q., & Kulkarni, C. (2020). “All rise for the AI director”: Eliciting possible futures of voice technology through story completion. *Proceedings of the 2020 ACM Designing Interactive Systems Conference*, 2051–2064. <https://doi.org/10.1145/3357236.3395479>
- Campbell, M., Hoane, A. J., & Hsu, F. H. (2002). Deep Blue. *Artificial Intelligence*, 134(1–2), 57–83. [https://doi.org/10.1016/S0004-3702\(01\)00129-1](https://doi.org/10.1016/S0004-3702(01)00129-1)
- Candello, H., Pinhanez, C., & Figueiredo, F. (2017). Typefaces and the perception of humanness in natural language chatbots. *Conference on Human Factors in Computing Systems - Proceedings*, 2017-May, 3476–3487. <https://doi.org/10.1145/3025453.3025919>
- Cesta, A., & D’alòisi, D. (1999). Mixed-Initiative Issues in an Agent-Based Meeting Scheduler. *Computational Models of Mixed-Initiative Interaction*, 229–262. https://doi.org/10.1007/978-94-017-1118-0_6
- Chandra, S., Srivastava, S. C., & Theng, Y. L. (2012). Cognitive absorption and trust for workplace collaboration in virtual worlds: An information processing decision making perspective. *Journal of the Association for Information Systems*, 13(10), 797–835. <https://doi.org/10.17705/1jais.00310>
- Chattaraman, V., Kwon, W. S., Gilbert, J. E., & Ross, K. (2019). Should AI-Based, conversational digital assistants employ social- or task-oriented interaction style?

- A task-competency and reciprocity perspective for older adults. *Computers in Human Behavior*, 90(August 2018), 315–330. <https://doi.org/10.1016/j.chb.2018.08.048>
- Chaves, A. P., & Gerosa, M. A. (2018). Single or multiple conversational agents? An interactional coherence comparison. *Conference on Human Factors in Computing Systems - Proceedings, 2018-April*. <https://doi.org/10.1145/3173574.3173765>
- Chen, S., Duckworth, K., & Chaiken, S. (1999). Motivated Heuristic and Systematic Processing. *Psychological Inquiry*, 10(1), 44–49. https://doi.org/10.1207/s15327965pli1001_6
- Cheng, Y., & Jiang, H. (2022). Customer–brand relationship in the era of artificial intelligence: understanding the role of chatbot marketing efforts. *Journal of Product and Brand Management*, 31(2), 252–264. <https://doi.org/10.1108/JPBM-05-2020-2907>
- Chernev, A. (2003). When More Is Less and Less Is More: The Role of Ideal Point Availability and Assortment in Consumer Choice. *Journal of Consumer Research*, 30(2), 170–183. <https://doi.org/10.1086/376808>
- Childers, T. L., Houston, M. J., & Heckler, S. E. (1985). Measurement of Individual Differences in Visual versus Verbal Information Processing. *Journal of Consumer Research*, 12(2), 125. <https://doi.org/10.1086/208501>
- Chin, H., Molefi, L. W., & Yi, M. Y. (2020). Empathy Is All You Need: How a Conversational Agent Should Respond to Verbal Abuse. *Conference on Human Factors in Computing Systems - Proceedings*, 1–13. <https://doi.org/10.1145/3313831.3376461>
- Chin, H., & Yi, M. Y. (2019). Should an agent be ignoring it? A study of verbal abuse types and conversational agents' response styles. *Conference on Human Factors in Computing Systems - Proceedings*, 1–6. <https://doi.org/10.1145/3290607.3312826>
- Cho, E. (2019). Hey Google, can i ask you something in private? The effects of modality and device in sensitive health information acquisition from voice assistants. *Conference on Human Factors in Computing Systems - Proceedings*. <https://doi.org/10.1145/3290605.3300488>
- Cho, J. (2006). The mechanism of trust and distrust formation and their relational outcomes. *Journal of Retailing*, 82(1), 25–35. <https://doi.org/10.1016/j.jretai.2005.11.002>

- Cialdini, R. B. (2009). *Influence: Science and Practice* (5th ed.). Pearson Education.
- Clark, H. H., & Brennan, S. E. (1991). Grounding in Communication. In *Perspectives on Socially Shared Cognition* (pp. 222–233). <https://doi.org/10.1037/10096-006>
- Clark, L., Munteanu, C., Wade, V., Cowan, B. R., Pantidi, N., Cooney, O., Doyle, P., Garaialde, D., Edwards, J., Spillane, B., Gilmartin, E., & Murad, C. (2019). What Makes a Good Conversation? *Proceedings of the 2019 CHI Conference Human Factors in Computing Systems*. <https://doi.org/10.1145/3290605.3300705>
- Clarke, S. G., & Haworth, J. T. (1994). ‘Flow’ experience in the daily lives of sixth-form college students. *British Journal of Psychology*, 85(4), 511–523. <https://doi.org/10.1111/j.2044-8295.1994.tb02538.x>
- Colby, K. M., Weber, S., & Hilf, F. D. (1971). Artificial Paranoia. *Artificial Intelligence*, 2(1), 1–25.
- Cooper, D. R., & Emory, C. W. (1991). *Business Research Methods* (4th ed.). Richard D Irwin.
- Cowan, B. R., Pantidi, N., Coyle, D., Morrissey, K., Clarke, P., Al-Shehri, S., Earley, D., & Bandeira, N. (2017). “What can i help you with?”: Infrequent users’ experiences of intelligent personal assistants. *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*. <https://doi.org/10.1145/3098279.3098539>
- Cowell, A. J., & Stanney, K. M. (2005). Manipulation of non-verbal interaction style and demographic embodiment to increase anthropomorphic computer character credibility. *International Journal of Human Computer Studies*, 62(2), 281–306. <https://doi.org/10.1016/j.ijhcs.2004.11.008>
- Crisp, R. J., & Hewstone, M. (2007). Multiple Social Categorization. *Advances in Experimental Social Psychology*, 39(06), 163–254. [https://doi.org/10.1016/S0065-2601\(06\)39004-1](https://doi.org/10.1016/S0065-2601(06)39004-1)
- Csikszentmihalyi, M. (1975). *Beyond boredom and anxiety: experiencing flow in work and play*. Jossey-Bass Publishers.
- Csikszentmihalyi, M. (1988). The flow experience and its significance for human psychologyo Title. *Optimal Experience: Psychological Studies of Flow in Consciousness*, 2, 15–35.

- Csikszentmihalyi, M., & LeFevre, J. (1989). Optimal Experience in Work and Leisure. *Journal of Personality and Social Psychology*, 56(5), 815–822. <https://doi.org/10.1037/0022-3514.56.5.815>
- Cummings, M. L., Buchin, M., Carrigan, G., & Donmez, B. (2010). Supporting intelligent and trustworthy maritime path planning decisions. *International Journal of Human Computer Studies*, 68(10), 616–626. <https://doi.org/10.1016/j.ijhcs.2010.05.002>
- Daft, R. L., & Lengel, R. H. (1986). Organizational Information Requirements, Media Richness and Structural Design. *Management Science*, 32(5), 554–571. <https://doi.org/10.1287/mnsc.32.5.554>
- Davenport, T., Guha, A., Grewal, D., & Bressgott, T. (2020). How artificial intelligence will change the future of marketing. *Journal of the Academy of Marketing Science*, 48, 24–42. <https://doi.org/10.1007/s11747-019-00696-0>
- de Bellis, E., & Venkataramani Johar, G. (2020). Autonomous Shopping Systems: Identifying and Overcoming Barriers to Consumer Adoption. *Journal of Retailing*, 96(1), 74–87. <https://doi.org/10.1016/j.jretai.2019.12.004>
- de Keyser, A., Köcher, S., Alkire (née Nasr), L., Verbeeck, C., & Kandampully, J. (2019). Frontline Service Technology infusion: conceptual archetypes and future research directions. *Journal of Service Management*, 30(1), 156–183. <https://doi.org/10.1108/JOSM-03-2018-0082>
- de Oliveira Santini, F., Ladeira, W. J., Pinto, D. C., Herter, M. M., Sampaio, C. H., & Babin, B. J. (2020). Customer engagement in social media: a framework and meta-analysis. *Journal of the Academy of Marketing Science*, 48(6), 1211–1228. <https://doi.org/10.1007/s11747-020-00731-5>
- de Visser, E. J., Monfort, S. S., McKendrick, R., Smith, M. A. B., McKnight, P. E., Krueger, F., & Parasuraman, R. (2016). Almost human: Anthropomorphism increases trust resilience in cognitive agents. *Journal of Experimental Psychology: Applied*, 22(3), 331–349. <https://doi.org/10.1037/xap0000092>
- Dehnert, M., & Mongeau, P. A. (2022). Persuasion in the Age of Artificial Intelligence (AI): Theories and Complications of AI-Based Persuasion. *Human Communication Research*.

- Dennis, A. R., Fuller, R. M., & Valacich, J. S. (2008). Media, tasks, and communication processes: A theory of media synchronicity. *MIS Quarterly*, 32(3), 575–600. <https://doi.org/10.2307/25148857>
- Dennis, A. R., & Kinney, S. T. (1998). Testing Media Richness Theory in the New Media: The Effects of Cues, Feedback, and Task Equivocality. *Information Systems Research*, 9(3), 256–274. <https://doi.org/10.1287/isre.9.3.256>
- Dennison, C. L. (2006). *The effect of gender stereotypes in languages on attitudes towards speakers* [Ph.D. Dissertation]. University of Pittsburgh.
- Diederich, S., Brendel, A. B., & Kolbe, L. M. (2020). Designing Anthropomorphic Enterprise Conversational Agents. *Business and Information Systems Engineering*. <https://doi.org/10.1007/s12599-020-00639-y>
- Diederich, S., Brendel, A. B., Lichtenberg, S., & Kolbe, L. (2019). DESIGN FOR FAST REQUEST FULFILLMENT OR NATURAL INTERACTION? INSIGHTS FROM AN EXPERIMENT WITH A CONVERSATIONAL AGENT. *27th European Conference on Information Systems (ECIS)*. https://aisel.aisnet.org/ecis2019_rp/20
- Diederich, S., Brendel, A. B., Morana, S., & Kolbe, L. (2022). On the Design of and Interaction with Conversational Agents : An Organizing and Assessing Review of Human-Computer Interaction Research. *Journal of the Association for Information Systems*, Forthcoming.
- D'Mello, S. K., Graesser, A., & King, B. (2010). Toward spoken human-computer tutorial dialogues. *Human-Computer Interaction*, 25(4), 289–323. <https://doi.org/10.1080/07370024.2010.499850>
- Dohle, S., & Montoya, A. K. (2017). The dark side of fluency: Fluent names increase drug dosing. *Journal of Experimental Psychology: Applied*, 23(3), 231–239. <https://doi.org/10.1037/xap0000131>
- Döring, N., & Bortz, J. (2016). *Forschungsmethoden und Evaluation*. Springer.
- Druga, S., Williams, R., Breazeal, C., & Resnick, M. (2017). “Hey Google is it OK if I eat you?” *Proceedings of the 2017 Conference on Interaction Design and Children*, 595–600. <https://doi.org/10.1145/3078072.3084330>
- DuBay, W. H. (2004). *The principles of readability: A brief introduction to readability research*. Impact Information.

- Dudley, H., Riesz, R. R., & Watkins, S. S. A. (1939). A synthetic speaker. *Journal of the Franklin Institute*, 227(6), 739–764.
- Eckes, T. (2008). Geschlechterstereotype: Von Rollen, Identitäten und Vorurteilen. In R. Becker & B. Kortendiek (Eds.), *Handbuch Frauen- und Geschlechterforschung. Theorie, Methoden & Empirie*. (pp. 171–182). VS.
- Eiband, M., Buschek, D., Kremer, A., & Hussmann, H. (2019). The Impact of Placebic Explanations on Trust in Intelligent Systems. *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, 1–6. <https://doi.org/10.1145/3290607.3312787>
- Elkins, A. C., & Derrick, D. C. (2013). The Sound of Trust: Voice as a Measurement of Trust During Interactions with Embodied Conversational Agents. *Group Decision and Negotiation*, 22(5), 897–913. <https://doi.org/10.1007/s10726-012-9339-x>
- Engel, C. (2022). *Cognitive Automation: Assessing Business Process Automation Potentials in Organizations Driven by Artificial Intelligence* [PhD Thesis]. University of St.Gallen.
- Engel, C., Ebel, P., & Leimeister, J. M. (2022). Cognitive automation. *Electronic Markets*. <https://doi.org/10.1007/s12525-021-00519-7>
- Engelhardt, S., Hansson, E., & Leite, I. (2017). Better faulty than sorry: Investigating social recovery strategies to minimize the impact of failure in human-robot interaction. *CEUR Workshop Proceedings*, 1943, 19–27.
- Epley, N., Waytz, A., Akalis, S., & Cacioppo, J. T. (2008). When we need a human: Motivational determinants of anthropomorphism. *Social Cognition*, 26(2), 143–155. <https://doi.org/10.1521/soco.2008.26.2.143>
- Eyssel, F., & Hegel, F. (2012). (S)he's Got the Look: Gender Stereotyping of Robots. *Journal of Applied Social Psychology*, 42(9), 2213–2230. <https://doi.org/10.1111/j.1559-1816.2012.00937.x>
- Fang, X., Singh, S., & Ahluwalia, R. (2007). An Examination of Different Explanations for. *Journal of Consumer Research*, 34(1), 97–103. <http://www.jstor.org/stable/10.1086/513050>
- Feine, J., Gnewuch, U., Morana, S., & Maedche, A. (2019). A Taxonomy of Social Cues for Conversational Agents. 132(June), 138–161. <https://doi.org/10.1016/j.ijhcs.2019.07.009>

- Feine, J., Gnewuch, U., Morana, S., & Maedche, A. (2020). Gender Bias in Chatbot Design. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. https://doi.org/10.1007/978-3-030-39540-7_6
- Fernández-Sabiote, E., & López-López, I. (2020). Discovering call interaction fluency: A way to improve experiences with call centres. *Service Science*, 12(1), 26–42. <https://doi.org/10.1287/SERV.2019.0251>
- Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, 82(6), 878–902. <https://doi.org/10.1037/0022-3514.82.6.878>
- Følstad, A., & Bae, P. (2020). Users ’ experiences with chatbots : findings from a questionnaire study. *Quality and User Experience*, 5(3). <https://doi.org/10.1007/s41233-020-00033-2>
- Følstad, A., & Brandtzaeg, P. B. (2017). Chatbots and the New World of HCI. *Interactions*, 24(4), 38–42. <https://doi.org/10.1145/3085558>
- Fornell, C., & Larcker, D. F. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research*, 18(1), 39. <https://doi.org/10.2307/3151312>
- Fransen, M. L., Fennis, B. M., & Pruyn, A. T. H. (2010). Matching communication modalities: The effects of modality congruence and processing style on brand evaluation and brand choice. *Communication Research*, 37(4), 576–598. <https://doi.org/10.1177/0093650210368251>
- Frantz, G. A., & Wiggins, R. H. (2013). Design case history: speak & spell learns to talk. *IEEE Spectrum*, 19(2), 45–49. <https://doi.org/10.1109/mspec.1982.6366797>
- Fuckner, M., Barthes, J. P., & Scalabrin, E. E. (2014). Using a personal assistant for exploiting service interfaces. *Proceedings of the 2014 IEEE 18th International Conference on Computer Supported Cooperative Work in Design*. <https://doi.org/10.1109/CSCWD.2014.6846822>
- Fuscaldo, D. (2019). *Bank Of America’s Virtual Assistant Now Has More Than 10 Million Users*. Forbes. <https://www.forbes.com/sites/donnafuscaldo/2019/12/11/bank-of-americas-virtual-assistant-now-has-more-than-10-million-users/?sh=1ceeb18ef69b>

- Gai, P. J., & Klesse, A. K. (2019). Making Recommendations More Effective Through Framings: Impacts of User- Versus Item-Based Framings on Recommendation Click-Throughs. *Journal of Marketing*, 83(6), 61–75. <https://doi.org/10.1177/0022242919873901>
- Gefen, D., & Straub, D. (2003). Managing User Trust in B2C e-Services. *E-Service Journal*, 2(2), 7. <https://doi.org/10.2979/esj.2003.2.2.7>
- Ghani, J. A., Supnick, R., & Rooney, P. (1991). The Experience of Flow in Computer-Mediated and in Face-to-Face Groups. *ICIS 1991 Proceedings*.
- Globerson, T. (1983). Mental capacity, mental effort, and cognitive style. *Developmental Review*, 3(3), 292–302.
- Gnewuch, U., Morana, S., Adam, M. T. P., & Maedche, A. (2018). Faster is not always better: Understanding the effect of dynamic response delays in human-chatbot interaction. *26th European Conference on Information Systems: Beyond Digitization - Facets of Socio-Technical Change, ECIS 2018*.
- Graf, L. K. M., & Landwehr, J. R. (2015). A Dual-Process Perspective on Fluency-Based Aesthetics: The Pleasure-Interest Model of Aesthetic Liking. *Personality and Social Psychology Review*, 19(4), 395–410. <https://doi.org/10.1177/1088868315574978>
- Graf, L. K. M., Mayer, S., & Landwehr, J. R. (2018). Measuring Processing Fluency: One versus Five Items. *Journal of Consumer Psychology*, 28(3), 393–411. <https://doi.org/10.1002/jcpy.1021>
- Gregor, S., & Benbasat, I. (2006). Explanations from Intelligent Systems: Theoretical Foundations and Implications for Practice. *MIS Quarterly*, 23(4), 497. <https://doi.org/10.2307/249487>
- Gregor, S., & Hevner, A. R. (2013). Positioning and Presenting Design Science Research for Maximum Impact. *MIS Quarterly*, 37(2), 337–355.
- Griffeth, E., & Simonite, T. (2019). *Facebook's Virtual Assistant M Is Dead. So Are Chatbots*.
- Guan, Z., Hou, F., Li, B., Phang, C. W., & Chong, A. Y. L. (2021). What influences the purchase of virtual gifts in live streaming in China? A cultural context-sensitive model. *Information Systems Journal*, 1–37. <https://doi.org/10.1111/isj.12367>

- Guiso, L. (2021). Trust and insurance. *Geneva Papers on Risk and Insurance: Issues and Practice*, 46(4), 509–512. <https://doi.org/10.1057/s41288-021-00241-7>
- Gustavsson, E. (2005). Virtual servants: Stereotyping female front-office employees on the Internet. *Gender, Work and Organization*, 12(5), 400–419. <https://doi.org/10.1111/j.1468-0432.2005.00281.x>
- Haeckel, S. H. (1998). About the nature and future of interactive marketing. *Journal of Interactive Marketing*, 12(1), 63–71.
- Hall, E. T. (1990). *The silent language*. Anchor books.
- Hammer, S., Wißner, M., & André, E. (2015). Trust-based decision-making for smart and adaptive environments. *User Modeling and User-Adapted Interaction*, 25(3), 267–293. <https://doi.org/10.1007/s11257-015-9160-8>
- Hannah, A., & Murachver, T. (1999). Gender and conversational style as predictors of conversational behavior. *Journal of Language and Social Psychology*, 18(2), 153–174. <https://doi.org/10.1177/0261927X99018002002>
- Harmeling, C. M., Moffett, J. W., Arnold, M. J., & Carlson, B. D. (2017). Toward a theory of customer engagement marketing. *Journal of the Academy of Marketing Science*, 45(3), 312–335. <https://doi.org/10.1007/s11747-016-0509-2>
- Hassenzahl, M., & Tractinsky, N. (2006). User experience - A research agenda. *Behaviour and Information Technology*, 25(2), 91–97. <https://doi.org/10.1080/01449290500330331>
- Hauswald, J., Laurenzano, M. A., Zhang, Y., Yang, H., Kang, Y., Li, C., Rovinski, A., Khurana, A., Dreslinski, R. G., Mudge, T., Petrucci, V., Tang, L., & Mars, J. (2016). Designing future warehouse-scale computers for sirius, an end-to-end voice and vision personal assistant. *ACM Transactions on Computer Systems*, 34(1). <https://doi.org/10.1145/2870631>
- Hayes, A. F. (2017). *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*. Guilford Press.
- Hayes, P., & Wagner, J. (2018). *Prepare for the voice revolution*. <https://www.pwc.com/us/en/services/consulting/library/consumer-intelligence-series/voice-assistants.html>

- Heeter, C. (1992). Being There: The Subjective Experience of Presence . *Presence: Teleoperators and Virtual Environments*, 1(2), 262–271. <https://doi.org/10.1162/pres.1992.1.2.262>
- Hefti, S. (2022). Hilfsbereit, verständnisvoll, bescheiden: Problematik und Folgen von weiblichen Voice Assistants. *HMD Praxis Der Wirtschaftsinformatik*, 59(2), 633–646. <https://doi.org/10.1365/s40702-022-00851-2>
- Heitmann, M., Landwehr, J. R., Schreiner, T. F., & van Heerde, H. J. (2020). Leveraging Brand Equity for Effective Visual Product Design. *Journal of Marketing Research*, 57(2), 257–277. <https://doi.org/10.1177/0022243720904004>
- Hess, T., Fuller, M., & Campbell, D. (2009). Designing interfaces with social presence: Using vividness and extraversion to create social recommendation agents. *Journal of the Association for Information Systems*, 10(12), 889–919. <https://doi.org/10.17705/1jais.00216>
- Hildebrand, C., & Bergner, A. (2020). Conversational robo advisors as surrogates of trust: onboarding experience, firm perception, and consumer financial decision making. *Journal of the Academy of Marketing Science*, 49, 659–676. <https://doi.org/10.1007/s11747-020-00753-z>
- Hildebrand, C., Efthymiou, F., Busquet, F., Hampton, W. H., Hoffman, D. L., & Novak, T. P. (2020). Voice analytics in business research: Conceptual foundations, acoustic feature extraction, and applications. *Journal of Business Research*, 121, 364–374. <https://doi.org/10.1016/j.jbusres.2020.09.020>
- Hill, J., Randolph Ford, W., & Farreras, I. G. (2015). Real conversations with artificial intelligence: A comparison between human-human online conversations and human-chatbot conversations. *Computers in Human Behavior*, 49, 245–250. <https://doi.org/10.1016/j.chb.2015.02.026>
- Hoffman, D. L., & Novak, T. P. (1996). Marketing in hypermedia computer-mediated environments: Conceptual foundations. *Journal of Marketing*, 60(3), 50–68. <https://doi.org/10.2307/1251841>
- Hoffman, D. L., & Novak, T. P. (2009). Flow Online: Lessons Learned and Future Prospects. *Journal of Interactive Marketing*, 23(1), 23–34. <https://doi.org/10.1016/j.intmar.2008.10.003>

- Hoffman, D. L., & Novak, T. P. (2018). Consumer and object experience in the internet of things: An assemblage theory approach. *Journal of Consumer Research*, 44(6), 1178–1204. <https://doi.org/10.1093/jcr/ucx105>
- Holbrook, M. B., & Gardner, M. P. (1993). An approach to investigating the emotional determinants of consumption durations: Why do people consume what they consume for as long as they consume it? *Journal of Consumer Psychology*, 2(2), 123–142. [https://doi.org/10.1016/S1057-7408\(08\)80021-6](https://doi.org/10.1016/S1057-7408(08)80021-6)
- Hollebeek, L. D., Sprott, D. E., & Brady, M. K. (2021). Rise of the Machines? Customer Engagement in Automated Service Interactions. *Journal of Service Research*, 24(1), 3–8. <https://doi.org/10.1177/1094670520975110>
- Holliday, D., Wilson, S., & Stumpf, S. (2016). User Trust in Intelligent Systems: A Journey over Time. *Proceedings of the 21st International Conference on Intelligent User Interfaces*, 164, 164–168. <https://doi.org/10.1016/j.ijid.2016.10.011>
- Hong, W., Thong, J. Y. L., & Tam, K. Y. (2004). The effects of information format and shopping task on consumers' online shopping behavior: A cognitive fit perspective. *Journal of Management Information Systems*, 21(3), 149–184. <https://doi.org/10.1080/07421222.2004.11045812>
- Huang, M. H., & Rust, R. T. (2018). Artificial Intelligence in Service. *Journal of Service Research*, 21(2), 155–172. <https://doi.org/10.1177/1094670517752459>
- Huang, M. H., & Rust, R. T. (2021). Engaged to a Robot? The Role of AI in Service. *Journal of Service Research*, 24(1), 30–41. <https://doi.org/10.1177/1094670520902266>
- Huber, B., McDuff, D., Brockett, C., Galley, M., & Dolan, B. (2018). Emotional Dialogue Generation using Image-Grounded Language Models. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. [https://doi.org/10.1016/0300-9572\(93\)90157-L](https://doi.org/10.1016/0300-9572(93)90157-L)
- Hulland, J., Baumgartner, H., & Smith, K. M. (2018). Marketing survey research best practices: evidence and recommendations from a review of JAMS articles. *Journal of the Academy of Marketing Science*, 46(1), 92–108. <https://doi.org/10.1007/s11747-017-0532-y>
- Hundertmark, S. (2021). *Telefonieren ist out - Chatten ist in*. Handelszeitung. <https://www.handelszeitung.ch/insurance/telefonieren-ist-out-chatten-ist-in>

- Isbister, K., & Nass, C. (2000). Consistency of personality in interactive characters: verbal cues, non-verbal cues, and user characteristics. *Interactional Human-Computer Studies*(2000), 53, 251–267. <https://doi.org/10.1006/ijhcs.2000.0368>
- Janssen, A., Passlick, J., Cordona, D. R., & Breitner, M. H. (2020). Virtual Assistance in Any Context: A Taxonomy of Design Elements for Domain-Specific Chatbots. *Business & Information Systems Engineering*, 62, 211–225. <https://doi.org/10.1007/s12599-020-00644-1>
- Jensen, M. L., Lowry, P. B., Burgoon, J. K., Jay, F., Jensen, M. L., Lowry, P. B., Burgoon, J. K., & Jr, J. F. N. (2010). Technology Dominance in Complex Decision Making : The Case of Aided Credibility Assessment. *Journal of Management Information Systems*, 27(1), 175–202. <https://doi.org/10.2753/MIS0742-1222270108>
- Jeong, Y., Kang, Y., & Lee, J. (2019). Exploring effects of conversational fillers on user perception of conversational agents. *Conference on Human Factors in Computing Systems - Proceedings*, 1–6. <https://doi.org/10.1145/3290607.3312913>
- Jeyaraj, A., Rottman, J. W., & Lacity, M. C. (2006). A review of the predictors, linkages, and biases in IT innovation adoption research. *Journal of Information Technology*, 21(1), 1–23. <https://doi.org/10.1057/palgrave.jit.2000056>
- Johnson, J. L., Lee, R. P. W., Saini, A., & Grohmann, B. (2003). Market-focused strategic flexibility: Conceptual advances and an integrative model. *Journal of the Academy of Marketing Science*, 31(1), 74–89. <https://doi.org/10.1177/0092070302238603>
- Junger, J., Pauly, K., Bröhr, S., Birkholz, P., Neuschaefer-Rube, C., Kohler, C., Schneider, F., Derntl, B., & Habel, U. (2013). Sex matters: Neural correlates of voice gender perception. *NeuroImage*, 79, 275–287. <https://doi.org/10.1016/j.neuroimage.2013.04.105>
- Jurafsky, D., & Martin, J. (2017). Dialog Systems and Chatbots. In *Speech and Language Processing* (pp. 418–440).
- Kahneman, D., & Frederick, S. (2007). Frames and brains: elicitation and control of response tendencies. *Trends in Cognitive Sciences*, 11(2), 45–46. <https://doi.org/10.1016/j.tics.2006.11.007>
- Kaneshige, T., & Hong, D. (2018). *Predictions 2019: This is the Year to Invest in Humans, as Backlash Against Chatbots and AI Begins*. Forrester. <https://go.forrester.com/blogs/predictions-2019-chatbots-and-ai-backlash/>

- Kang, J., & Wei, L. (2018). "Give Me the support I want!": The effect of matching an embodied conversational agent's social support to users' social support needs in fostering positive user-agent interaction. *HAI 2018 - Proceedings of the 6th International Conference on Human-Agent Interaction*, 106–113. <https://doi.org/10.1145/3284432.3284462>
- Karnik, A. (2019). *The Who, When And Why Of Delivering Personalized Product Recommendations*. Forbes. <https://www.forbes.com/sites/forbescommunicationscouncil/2019/04/22/the-who-when-and-why-of-delivering-personalized-product-recommendations/?sh=87a0f3e1b925>
- Kato, I., Ohteru, S., Kobayashi, H., Shirai, K., & Uchiyama, A. (1972). Information-Power Machine with Senses and Limbs. In *On Theory and Practice of Robots and Manipulators*. International Centre for Mechanical Sciences. Springer.
- Khawaja, M. A., Chen, F., & Marcus, N. (2010). Using Language Complexity to Measure Cognitive Load for Adaptive Interaction Design. *International Conference on Intelligent User Interfaces, Proceedings IUI*, 333–336. <https://doi.org/10.1145/1719970.1720024>
- Kim, A., Cho, M., Ahn, J., & Sung, Y. (2019). Effects of Gender and Relationship Type on the Response to Artificial Intelligence. *Cyberpsychology, Behavior, and Social Networking*, 22(4), 249–253. <https://doi.org/10.1089/cyber.2018.0581>
- Kim, D., & Ko, Y. J. (2019). The impact of virtual reality (VR) technology on sport spectators' flow experience and satisfaction. *Computers in Human Behavior*, 93, 346–356. <https://doi.org/10.1016/j.chb.2018.12.040>
- Kim, S., Lee, J., & Gweon, G. (2019, May 2). Comparing data from chatbot and web surveys effects of platform and conversational style on survey response quality. *Conference on Human Factors in Computing Systems - Proceedings*. <https://doi.org/10.1145/3290605.3300316>
- Kim, S., Row, Y. K., & Nam, T. J. (2018). Thermal interaction with a voice-based intelligent agent. *Conference on Human Factors in Computing Systems - Proceedings, 2018-April*. <https://doi.org/10.1145/3170427.3188656>
- King, D., Auschaitrakul, S., & Lin, C. W. J. (2021). Search modality effects: merely changing product search modality alters purchase intentions. *Journal of the Academy of Marketing Science*. <https://doi.org/10.1007/s11747-021-00820-z>

- Kinsella, B. (2020). *Privacy Concerns Rise Significantly as 1-in-3 Consumers Cite it as Reason to Avoid Smart Speakers*. <https://voicebot.ai/2020/05/11/privacy-concerns-rise-significantly-as-1-in-3-consumers-cite-it-as-reason-to-avoid-smart-speakers/>
- Kirk, R. E. (1982). Procedures for testing hypotheses about means: Rules governing the choice of error term for simple effect means and simple main effects. In *Experimental design: Procedures for the behavioral sciences* (2nd ed., pp. 509–510). Brooks/Cole.
- Klesse, A. K., Levav, J., & Goukens, C. (2015). The effect of preference expression modality on self-control. *Journal of Consumer Research*, 42(4), 535–550. <https://doi.org/10.1093/jcr/ucv043>
- Knapp, M. L., Hall, J. A., & Horgan, T. G. (2013). *Nonverbal communication in human interaction*. Cengage Learning.
- Knote, R., Janson, A., Eigenbrod, L., & Söllner, M. (2018). The what and how of smart personal assistants: Principles and application domains for IS research. *MKWI 2018 - Multikonferenz Wirtschaftsinformatik, 2018-March*, 1083–1094.
- Knote, R., Janson, A., Söllner, M., & Leimeister, J. M. (2019). Classifying Smart Personal Assistants: An Empirical Cluster Analysis. *Proceedings of the 52nd Hawaii International Conference on System Sciences*, 2024–2033. <https://doi.org/10.24251/hicss.2019.245>
- Knote, R., Janson, A., Söllner, M., & Leimeister, J. M. (2021). Value Co-Creation in Smart Services: A Functional Affordances Perspective on Smart Personal Assistants. *Journal of the Association for Information Systems*, 22(2), 418–458.
- Kobori, T., Nakano, M., & Nakamura, T. (2016). Small Talk Improves User Impressions of Interview Dialogue Systems. *SIGDIAL 2016 - 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue, Proceedings of the Conference, September*, 370–380. <https://doi.org/10.18653/v1/w16-3646>
- Koenig, A. M., Eagly, A. H., Mitchell, A. A., & Ristikari, T. (2011). Are leader stereotypes masculine? A meta-analysis of three research paradigms. *Psychological Bulletin*, 137(4), 616–642. <https://doi.org/10.1037/a0023557>
- Komiak, S. X., & Benbasat, I. (2003). Understanding Customer Trust in Agent-Mediated Electronic Commerce, Web-Mediated Electronic Commerce, and Traditional Commerce. *Information Technology and Management*, 5, 181–207. <https://doi.org/10.1023/b:item.0000008081.55563.d4>

- Kontogiorgos, D. (2022). *Mutual Understanding in Situated Interactions with Conversational User Interfaces: Theory, Studies, and Computation* [PhD Thesis]. KTH Royal Institute of Technology.
- Kostyk, A., Niculescu, M., & Leonhardt, J. M. (2017). Less is more: Online consumer ratings' format affects purchase intentions and processing. *Journal of Consumer Behaviour*, 16(5), 434–441. <https://doi.org/10.1002/cb.1643>
- Koulouri, T., Lauria, S., & Macredie, R. D. (2016). Do (and Say) as I Say: Linguistic adaptation in human–computer dialogs. *Human-Computer Interaction*, 31(1), 59–95. <https://doi.org/10.1080/07370024.2014.934180>
- Krämer, N. C., Iurgel, I. A., & Bente, G. (2005). Emotion and motivation in embodied conversational agentse. *Proceedings of the Symposium on Agents That Want and Like: Motivational and Emotional Roots of Cognition and Action*.
- Krämer, N. C., Kopp, S., Becker-Asano, C., & Sommer, N. (2013). Smile and the world will smile with you - The effects of a virtual agent's smile on users' evaluation and behavior. *International Journal of Human Computer Studies*, 71(3), 335–349. <https://doi.org/10.1016/j.ijhcs.2012.09.006>
- Kroque, K. (2017). Artificial Intelligence is Here to Stay, but Constumer Trust is a Must for AI in Business. *Forbes*.
- Lacity, M. C., Khan, S., Yan, A., & Willcocks, L. P. (2010). A review of the IT outsourcing empirical literature and future research directions. *Journal of Information Technology*, 25(4), 395–433. <https://doi.org/10.1057/jit.2010.21>
- Laricchia, F. (2022a). <https://www.statista.com/statistics/917900/selected-countries-amazon-alexa-skill-count/>. Statista. <https://www.statista.com/statistics/917900/selected-countries-amazon-alexa-skill-count/>
- Laricchia, F. (2022b). *Number of digital voice assistants in use worldwide from 2019 to 2024 (in billions)**. Statista. <https://www.statista.com/statistics/973815/worldwide-digital-voice-assistant-in-use/>
- Larivière, B., Bowen, D., Andreassen, T. W., Kunz, W., Sirianni, N. J., Voss, C., Wunderlich, N. v., & de Keyser, A. (2017). “Service Encounter 2.0”: An investigation into the roles of technology, employees and customers. *Journal of Business Research*, 79, 238–246. <https://doi.org/10.1016/j.jbusres.2017.03.008>

- Laumer, S., Maier, C., & Gubler, F. T. (2019). Chatbot Acceptance in Healthcare: Explaining User Adoption of Conversational Agents for Disease Diagnosis. *Proceedings of the 27th European Conference on Information Systems*. https://aisel.aisnet.org/ecis2019_rp/88
- Law, E. L. C., Roto, V., Hassenzahl, M., Vermeeren, A. P. O. S., & Kort, J. (2009). Understanding, scoping and defining user experience: A survey approach. *Conference on Human Factors in Computing Systems - Proceedings, April*, 719–728. <https://doi.org/10.1145/1518701.1518813>
- le Bigot, L., Rouet, J. F., & Jamet, E. (2007). Effects of speech- and text-based interaction modes in natural language human-computer dialogue. *Human Factors*, 49(6), 1045–1053. <https://doi.org/10.1518/001872007X249901>
- Leathers, D. G., & Eaves, M. (2015). *Successful nonverbal communication: Principles and application*. Routledge.
- Lee, E. J., Nass, C., & Brave, S. (2000). Can computer-generated speech have gender? An experimental test of gender stereotype. *CHI'00 Extended Abstracts on Human Factors in Computing Systems, April*, 289–290. <https://doi.org/10.1145/633292.633461>
- Lee, G. G., & Lin, H. F. (2005). Customer perceptions of e-service quality in online shopping. *International Journal of Retail and Distribution Management*, 33(2), 161–176. <https://doi.org/10.1108/09590550510581485>
- Lee, J. D., & See, K. A. (2004). Trust in Automation: Designing for Appropriate Reliance. *Human Factors*, 46(1), 50–80.
- Lee, J. G., Kim, K. J., Lee, S., & Shin, D. H. (2015). Can Autonomous Vehicles Be Safe and Trustworthy? Effects of Appearance and Autonomy of Unmanned Driving Systems. *International Journal of Human-Computer Interaction*, 31(10), 682–691. <https://doi.org/10.1080/10447318.2015.1070547>
- Lee, S., Lee, S., & Kim, S. (2019, May 2). “What does your agent look like?” A drawing study to understand users’ perceived persona of conversational agent. *Conference on Human Factors in Computing Systems - Proceedings*. <https://doi.org/10.1145/3290607.3312796>
- Lee, S., Subramani, M., Kiesler, S., Walker, J. H., & Waters, K. (1996). When the interface is a face. *Human-Computer Interaction*, 11(2), 97–124. https://doi.org/10.1207/s15327051hci1102_1

- Lee, S. Y., & 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(May 2016), 95–105. <https://doi.org/10.1016/j.ijhcs.2017.02.005>
- Lehrer, C., & Trenz, M. (2022). Omnichannel Business. *Electronic Markets*, 0123456789. <https://doi.org/10.1007/s12525-021-00511-1>
- Lemon, K. N., & Verhoef, P. C. (2016). Understanding customer experience throughout the customer journey. *Journal of Marketing*, 80(6), 69–96. <https://doi.org/10.1509/jm.15.0420>
- Levinson, S. C. (2016). Turn-taking in Human Communication - Origins and Implications for Language Processing. *Trends in Cognitive Sciences*, 20(1), 6–14. <https://doi.org/10.1016/j.tics.2015.10.010>
- Li, N., & Zhang, P. (2005). The Intellectual Development of Human-Computer Interaction Research: A Critical Assessment of the MIS Literature (1990-2002). *Journal of the Association for Information Systems*, 6(11), 227–292.
- Liao, M., & Sundar, S. S. (2021). When E-Commerce Personalization Systems Show and Tell: Investigating the Relative Persuasive Appeal of Content-Based versus Collaborative Filtering. *Journal of Advertising*. <https://doi.org/10.1080/00913367.2021.1887013>
- Liao, Q. V., Hussain, M. M. U., Chandar, P., Davis, M., Khazaen, Y., Crasso, M. P., Wang, D., Muller, M., Shami, N. S., & Geyer, W. (2018). All work and no play? Conversations with a question-and-answer chatbot in the wild. *Conference on Human Factors in Computing Systems - Proceedings, 2018-April*. <https://doi.org/10.1145/3173574.3173577>
- Loideain, N. N., & Adams, R. (2020). From Alexa to Siri and the GDPR: The gendering of Virtual Personal Assistants and the role of Data Protection Impact Assessments. *Computer Law and Security Review*, 36, 1–14. <https://doi.org/10.1016/j.clsr.2019.105366>
- Lovett, L. (2019). *Consumers interested in voice tech for health, adoption remains low, survey reports*. <https://www.mobihealthnews.com/news/north-america/consumers-interested-voice-tech-health-adoption-remains-low-survey-reports>

- Lowry, P. B., D'Arcy, J., Hammer, B., & Moody, G. D. (2016). "Cargo Cult" science in traditional organization and information systems survey research: A case for using nontraditional methods of data collection, including Mechanical Turk and online panels. *Journal of Strategic Information Systems*, 25(3), 232–240. <https://doi.org/10.1016/j.jsis.2016.06.002>
- Maedche, A., Legner, C., Benlian, A., Berger, B., Gimpel, H., Hess, T., Hinz, O., Morana, S., & Söllner, M. (2019). AI-Based Digital Assistants. *Business & Information Systems Engineering*, 61(4), 535–544. <https://doi.org/10.1007/s12599-019-00600-8>
- Maedche, A., Morana, S., Schacht, S., Werth, D., & Krumeich, J. (2016). Advanced user assistance systems. *Business and Information Systems Engineering*, 58(5), 367–370. <https://doi.org/10.1007/s12599-016-0444-2>
- Mathwick, C., & Rigdon, E. (2004). Play, flow, and the online search experience. *Journal of Consumer Research*, 31(2), 324–332. <https://doi.org/10.1086/422111>
- Mauldin, M. L. (1994). ChatterBots, tinyMuds, and the turing test entering the loebner prize competition. *Proceedings of the National Conference on Artificial Intelligence*, 1, 16–21.
- Mauze, B., Venkatesan, C., Cahndrashekar, A., & Agarwal, C. (2018). *Voice assistants - the new glass!* <https://www.capgemini.com/ch-en/wp-content/uploads/sites/43/2018/10/Voice-Assistants.pdf>
- Mayer, R. C., Davis, J. H., & Schoorman, D. F. (1995). An Integrative Model of Organizational Trust. *The Academy of Management Review*, 20(3), 709–734.
- Mayer, R. E., Johnson, W. L., Shaw, E., & Sandhu, S. (2006). Constructing computer-based tutors that are socially sensitive: Politeness in educational software. *International Journal of Human Computer Studies*, 64(1), 36–42. <https://doi.org/10.1016/j.ijhcs.2005.07.001>
- Mc Knight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing And Validating Trust Measure for E-Commerce: An Integrative Typology. *Information Systems Research*, 13(3), 334–359. <https://www-1jstor-1org-10011f5ps07a5.emedia1.bsb-muenchen.de/stable/pdf/23015741.pdf?refreqid=excel-sior%3Aa1eb16b22a2ceb93cf3bf1392d1f106a%0Ahttps://search.proquest.com/docview/208146326/fulltextPDF/8E5767BEFF7D410DPQ/1?accountid=11107>

- McAleer, P., Todorov, A., & Belin, P. (2014). How do you say “hello”? Personality impressions from brief novel voices. *PLoS ONE*, 9(3), 1–9. <https://doi.org/10.1371/journal.pone.0090779>
- McBreen, H. M., & Jack, M. A. (2001). Evaluating humanoid synthetic agents in e-retail applications. *IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans*, 31(5), 394–405. <https://doi.org/10.1109/3468.952714>
- McDowell, W. C., Wilson, R. C., & Kile, C. O. (2016). An examination of retail website design and conversion rate. *Journal of Business Research*, 69(11), 4837–4842. <https://doi.org/10.1016/j.jbusres.2016.04.040>
- Mcgrath, J. E. (1995). METHODOLOGY MATTERS: DOING RESEARCH IN THE BEHAVIORAL and SOCIAL SCIENCES. In *Readings in Human-Computer Interaction* (pp. 152–169). Morgan Kaufmann.
- McKinney, V., Yoon, K., & Zahedi, F. (2002). The measurement of Web-customer satisfaction: An expectation and disconfirmation approach. *Information Systems Research*, 13(3), 296–315. <https://doi.org/10.1287/isre.13.3.296.76>
- McKnight, D. H., Carter, M., Thatcher, J. B., & Clay, P. F. (2011). Trust in a specific technology: An investigation of its components and measures. *ACM Transactions on Management Information Systems*, 2(2). <https://doi.org/10.1145/1985347.1985353>
- McKnight, D. H., Liu, P., & Pentland, B. T. (2020). Trust Change in Information Technology Products. *Journal of Management Information Systems*, 37(4), 1015–1046. <https://doi.org/10.1080/07421222.2020.1831772>
- McLean, G., & Osei-Frimpong, K. (2019). Hey Alexa ... examine the variables influencing the use of artificial intelligent in-home voice assistants. *Computers in Human Behavior*, 99, 28–37. <https://doi.org/10.1016/j.chb.2019.05.009>
- McTear, M., Callejas, Z., & Griol, D. (2016). The conversational interface: Talking to smart devices. In *The Conversational Interface: Talking to Smart Devices*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-32967-3>
- Melumad, S., & Meyer, R. (2020). Full Disclosure: How Smartphones Enhance Consumer Self-Disclosure. *Journal of Marketing*, 84(3), 28–45. <https://doi.org/10.1177/0022242920912732>

- Mende, M., Scott, M. L., van Doorn, J., Grewal, D., & Shanks, I. (2019). Service Robots Rising: How Humanoid Robots Influence Service Experiences and Elicit Compensatory Consumer Responses. *Journal of Marketing Research*, 56(4), 535–556. <https://doi.org/10.1177/0022243718822827>
- Mesbah, N., Olt, C. M., Tauchert, C., & Buxmann, P. (2019). Promoting Trust in AI-based Expert Systems. *Proceedings of the 25th Americas Conference on Information Systems*.
- Meza Martínez, M. A., Nadj, M., & Maedche, A. (2019). Towards an Integrative Theoretical Framework of Interactive Machine Learning Systems. *Proceedings of the 27th European Conference on Information Systems*.
- Miao, F., Kozlenkova, I. v., Wang, H., Xie, T., & Palmatier, R. W. (2022). An Emerging Theory of Avatar Marketing. *Journal of Marketing*, 68(1), 67–90. <https://doi.org/10.1177/0022242921996646>
- Miehle, J., Minker, W., & Ultes, S. (2018). Exploring the impact of elaborateness and indirectness on user satisfaction in a spoken dialogue system. *UMAP 2018 - Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization*, 165–172. <https://doi.org/10.1145/3213586.3226213>
- Mitchell, W. J., Ho, C. C., Patel, H., & MacDorman, K. F. (2011). Does social desirability bias favor humans? Explicit-implicit evaluations of synthesized speech support a new HCI model of impression management. *Computers in Human Behavior*, 27(1), 402–412. <https://doi.org/10.1016/j.chb.2010.09.002>
- Moffett, J. W., Folse, J. A. G., & Palmatier, R. W. (2021). A theory of multiformat communication: mechanisms, dynamics, and strategies. *Journal of the Academy of Marketing Science*, 49, 441–461. <https://doi.org/10.1007/s11747-020-00750-2>
- Mogilner, C., Shiv, B., & Iyengar, S. S. (2013). Eternal quest for the best: Sequential (vs. Simultaneous) option presentation undermines choice commitment. *Journal of Consumer Research*, 39(6), 1300–1312. <https://doi.org/10.1086/668534>
- Monti, M., Pelligra, V., Martignon, L., & Berg, N. (2014). Retail investors and financial advisors: New evidence on trust and advice taking heuristics. *Journal of Business Research*, 67(8), 1749–1757. <https://doi.org/10.1016/j.jbusres.2014.02.022>
- Moon, Y. (2000). Intimate exchanges: Using computers to elicit self-disclosure from consumers. *Journal of Consumer Research*, 26(4), 323–339. <https://doi.org/10.1086/209566>

- Morgan, C., Fajardo, T. M., & Townsend, C. (2021). Show it or say it: how brand familiarity influences the effectiveness of image-based versus text-based logos. *Journal of the Academy of Marketing Science*, 49(3), 566–583. <https://doi.org/10.1007/s11747-020-00760-0>
- Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley. *IEEE Robotics and Automation Magazine*, 19(2), 98–100. <https://doi.org/10.1109/MRA.2012.2192811>
- Mosteller, J., Donthu, N., & Eroglu, S. (2014). The fluent online shopping experience. *Journal of Business Research*, 67(11), 2486–2493. <https://doi.org/10.1016/j.jbusres.2014.03.009>
- Mu, J., & Sarkar, A. (2019). Do We Need Natural Language? *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–6. <https://doi.org/10.1145/3290607.3312975>
- Müller-Bloch, C., & Kranz, J. (2015). A framework for rigorously identifying research gaps in qualitative literature reviews. *International Conference on Information Systems (ICIS)*, 1–19.
- Muralidharan, L., de Visser, E. J., & Parasuraman, R. (2014). The effects of pitch contour and flanging on trust in speaking cognitive agents. *Extended Abstracts of the 2014 CHI Conference Human Factors in Computing Systems - CHI '14*, 2167–2172. <https://doi.org/10.1145/2559206.2581231>
- Myers, C., Furqan, A., Nebolsky, J., Caro, K., & Zhu, J. (2018). Patterns for How Users Overcome Obstacles in Voice User Interfaces. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. <http://dl.acm.org/citation.cfm?doid=3173574.3173580>
<http://doi.acm.org/10.1145/3173574.3173580>
- Nass, C., & Lee, K. M. (2001). Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, and consistency-attraction. *Journal of Experimental Psychology: Applied*, 7(3), 171–181. <https://doi.org/10.1037/1076-898X.7.3.171>
- Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, 56(1), 81–103. <https://doi.org/10.1111/0022-4537.00153>

- Nass, C., Moon, Y., & Green, N. (1997). Are machines gender neutral? Gender-stereotypic responses to computers with voices. *Journal of Applied Social Psychology*, 27(10), 864–876. <https://doi.org/10.1111/j.1559-1816.1997.tb00275.x>
- Nass, C., Steuer, J., & Tauber, E. R. (1994). Computer are social actors. *Conference on Human Factors in Computing Systems - Proceedings*, 72–78.
- Nielsen, F. Å. (2011). A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. *CEUR Workshop Proceedings*, 93–98. <http://arxiv.org/abs/1103.2903>
- Nordheim, C. B., Følstad, A., & Bjørkli, C. A. (2019). An Initial Model of Trust in Chatbots for Customer Service—Findings from a Questionnaire Study. *Interacting with Computers*, 31(3), 317–335. <https://doi.org/10.1093/iwc/iwz022>
- Novak, T. P., & Hoffman, D. L. (2018). Relationship journeys in the internet of things: a new framework for understanding interactions between consumers and smart objects. *Journal of the Academy of Marketing Science*, 47(2), 216–237. <https://doi.org/10.1007/s11747-018-0608-3>
- Novak, T. P., Hoffman, D. L., Yung, Y. F., Novak, T. P., & Hoffman, D. L. (2000). Measuring the customer experience in online environments: A structural modeling approach. *Marketing Science*, 19(1), 22–42. <https://doi.org/10.1287/mksc.19.1.22.15184>
- Novemsky, N., Dhar, R., Schwarz, N., & Simonson, I. (2007). Preference fluency in choice. *Journal of Marketing Research*, 44(3), 347–356. <https://doi.org/10.1509/jmkr.44.3.347>
- Nowak, K. L., & Biocca, F. (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), 481–494. <https://doi.org/10.1162/105474603322761289>
- Nunamaker, J. F., Derrick, D. C., Elkins, A. C., Burgoon, J. K., & Patton, M. W. (2011). Embodied Conversational Agent-Based Kiosk for Automated Interviewing. *Journal of Management Information Systems*, 28(1), 17–48. <https://doi.org/10.2753/mis0742-1222280102>
- O'Connor, J. J. M., & Barclay, P. (2017). The influence of voice pitch on perceptions of trustworthiness across social contexts. *Evolution and Human Behavior*, 38(4), 506–512. <https://doi.org/10.1016/j.evolhumbehav.2017.03.001>

- Oord, A. van den, Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A., & Kavukcuoglu, K. (2016). *WaveNet: A Generative Model for Raw Audio*. 1–15. <http://arxiv.org/abs/1609.03499>
- Pagani, M., Racat, M., & Hofacker, C. F. (2019). Adding Voice to the Omnichannel and How that Affects Brand Trust. *Journal of Interactive Marketing*, 48, 89–105. <https://doi.org/10.1016/j.intmar.2019.05.002>
- Pak, R., McLaughlin, A. C., & Bass, B. (2014). A multi-level analysis of the effects of age and gender stereotypes on trust in anthropomorphic technology by younger and older adults. *Ergonomics*, 57(9), 1277–1289. <https://doi.org/10.1080/00140139.2014.928750>
- Park, E. K., & Sundar, S. S. (2015). Can synchronicity and visual modality enhance social presence in mobile messaging? *Computers in Human Behavior*, 45, 121–128. <https://doi.org/10.1016/j.chb.2014.12.001>
- Parviainen, E., & Søndergaard, M. L. J. (2020). Experiential Qualities of Whispering with Voice Assistants. *Proceedings of the 2020 Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/3313831.3376187>
- Pecune, F., Chen, J., Matsuyama, Y., & Cassell, J. (2018). Field Trial Analysis of Socially Aware Robot Assistant. *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*, 9, 1241–1249.
- Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, 70, 153–163. <https://doi.org/10.1016/j.jesp.2017.01.006>
- Peer, E., Rothschild, D., Evernden, Z., Gordon, A., & Damer, E. (2021). Data Quality of Platforms and Panels for Online Behavioral Research Data Quality of Platforms and Panels for Online Behavioral Research. *Social Science Research Network*, 1–46.
- Perez, S. (2017). *Starbucks unveils a virtual assistant that takes your order via messaging or voice*. TechCrunch. <https://techcrunch.com/2017/01/30/starbucks-unveils-a-virtual-assistant-that-takes-your-order-via-messaging-or-voice/>
- Pfeuffer, N., Adam, M., Toutaoui, J., Hinz, O., & Benlian, A. (2019). Mr. and Mrs. Conversational Agent - Gender Stereotyping in Judge-Advisor Systems and the Role of Egocentric Bias. *ICIS 2019 Proceedings*.

- https://aisel.aisnet.org/icis2019/human_computer_interact/human_computer_interact/2
- Pfeuffer, N., Benlian, A., Gimpel, H., & Hinz, O. (2019). Anthropomorphic Information Systems. *Business & Information Systems Engineering*, 61(4), 523–533. <https://doi.org/10.1007/s12599-019-00599-y>
- Pogacar, R., Shrum, L. J., & Lowrey, T. M. (2018). The Effects of Linguistic Devices on Consumer Information Processing and Persuasion: A Language Complexity $\times$ Processing Mode Framework. *Journal of Consumer Psychology*, 28(4), 689–711. <https://doi.org/10.1002/jcpy.1052>
- Press, G. (2019). *AI Stats News: 62% Of US Consumers Like Using Chatbots To Interact With Businesses*. Forbes. <https://www.forbes.com/sites/gilpress/2019/10/25/ai-stats-news-us-consumers-interest-in-using-chatbots-to-interact-with-businesses-rise-to-62/?sh=467d5e5c3974>
- Purington, A., Taft, J. G., Sannon, S., Bazarova, N. N., & Taylor, S. H. (2017). “Alexa is my new BFF”: Social Roles, User Satisfaction, and Personification of the Amazon Echo. *Extended Abstracts of the 2017 CHI Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/3027063.3053246>
- Q. (2019). *The First Genderless Voice*. GenderLess Voice. <https://www.genderlessvoice.com/>
- Qiu, L., & Benbasat, I. (2009). Evaluating anthropomorphic product recommendation agents: A social relationship perspective to designing information systems. *Journal of Management Information Systems*, 25(4), 145–182. <https://doi.org/10.2753/MIS0742-1222250405>
- Qiu, L., & Benbasat, I. (2010). A study of demographic embodiments of product recommendation agents in electronic commerce. *International Journal of Human Computer Studies*, 68(10), 669–688. <https://doi.org/10.1016/j.ijhcs.2010.05.005>
- Qiu, L., & Benbesat, I. (2005). Online consumer trust and live help interfaces: The effects of text-to-speech voice and three-dimensional avatars. *International Journal of Human-Computer Interaction*, 19(1), 75–94.
- Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J. F., Breazeal, C., Crandall, J. W., Christakis, N. A., Couzin, I. D., Jackson, M. O., Jennings, N. R., Kamar, E., Kloumann, I. M., Larochelle, H., Lazer, D., McElreath, R., Mislove, A.,

- Parkes, D. C., Pentland, A. 'Sandy,' ... Wellman, M. (2019). Machine behaviour. *Nature*, 568(7753), 477–486. <https://doi.org/10.1038/s41586-019-1138-y>
- Rai, A. (2017). Avoiding type III errors: Formulating is research problems that matter. *MIS Quarterly: Management Information Systems*, 41(2), iii–vii.
- Recker, J. (2013). *Scientific Research in Information Systems: A Binner's Guide*. Springer.
- Redeker, G. (1984). On Differences Between Spoken and Written Language. *Discourse Processes*, 7(1), 43–55. <https://doi.org/10.1080/01638538409544580>
- Reinkemeier, F., & Gnewuch, U. (2022). Match or Mismatch? How Matching Personality and Gender between Voice Assistants and Users Affects Trust in Voice Commerce. *Proceedings of the 55th Hawaii International Conference on System Sciences, January*. <https://doi.org/10.24251/hicss.2022.528>
- Rhee, C. E., & Choi, J. (2020). Effects of personalization and social role in voice shopping: An experimental study on product recommendation by a conversational voice agent. *Computers in Human Behavior*, 109(March), 106359. <https://doi.org/10.1016/j.chb.2020.106359>
- Riccardi, G. (2014). Towards healthcare personal agents. *RFMIR 2014 - Proceedings of the 2014 ACM Roadmapping the Future of Multimodal Interaction Research Including Business Opportunities and Challenges*, 53–56. <https://doi.org/10.1145/2666253.2666266>
- Riedl, R., Mohr, P. N. C., Kenning, P. H., Davis, F. D., & Heekeren, H. R. (2014). Trusting Humans and Avatars: A Brain Imaging Study Based on Evolution Theory. *Journal of Management Information Systems*, 30(4), 83–114. <https://doi.org/10.2753/mis0742-1222300404>
- Riikkinen, M., Saarijärvi, H., Sarlin, P., & Lähteenmäki, I. (2018). Using artificial intelligence to create value in insurance. *International Journal of Bank Marketing*, 36(6), 1145–1168. <https://doi.org/10.1108/IJBM-01-2017-0015>
- Rockmann, K. W., & Northcraft, G. B. (2008). To be or not to be trusted: The influence of media richness on defection and deception. *Organizational Behavior and Human Decision Processes*, 107(2), 106–122. <https://doi.org/10.1016/j.obhdp.2008.02.002>
- Rousseau, D. M., Burt, R. S., & Camerer, C. (1998). Not so Different after All : A Cross-Discipline View of Trust. *The Academy of Management Review*, 23(3), 393–404.

- Rubin, D. L., Hafer, T., & Arata, K. (2000). Reading and listening to oral-based versus literate-based discourse. *Communication Education*, 49(2), 121–133. <https://doi.org/10.1080/03634520009379200>
- Rzepka, C., & Berger, B. (2018). User Interaction with AI-enabled Systems: A systematic review of IS research. *Proceedings of the 39th International Conference on Information Systems*.
- Rzepka, C., Berger, B., & Hess, T. (2021). Voice Assistant vs . Chatbot – Examining the Fit Between Conversational Agents’ Interaction Modalities and Information Search Tasks. *Information Systems Frontiers*, Forthcoming. <https://doi.org/10.1007/s10796-021-10226-5>
- Schaffer, J., O’donovan, J., Michaelis, J., Raglin, A., & Höllerer, T. (2019). I Can Do Better Than Your AI: Expertise and Explanations ACM Reference Format. *Proceedings of the 24th ACM International Conference on Intelligent User Interfaces*, 240–251. <https://doi.org/10.1145/3301275.3302308>
- Schanke, S., Burtch, G., & Ray, G. (2021). Estimating the impact of “humanizing” customer service chatbots. *Information Systems Research*, 32(3), 736–751. <https://doi.org/10.1287/ISRE.2021.1015>
- Schlosser, A. E., White, T. B., & Lloyd, S. M. (2006). Converting web site visitors into buyers: How web site investment increases consumer trusting beliefs and online purchase intentions. *Journal of Marketing*, 70(2), 133–148. <https://doi.org/10.1509/jmkg.70.2.133>
- Schmitt, A., Wambsganss, T., Söllner, M., & Janson, A. (2021). Towards a Trust Reliance Paradox? Exploring the Gap Between Perceived Trust in and Reliance on Algorithmic Advice. *Forty-Second International Conference on Information Systems, October*.
- Schmitt, A., Zierau, N., Janson, A., & Leimeister, J. M. (2021). Voice as a Contemporary Frontier of Interaction Design. *European Conference of Information Systems (ECIS), June*.
- Schniter, E., & Shields, T. W. (2020). Gender, Stereotypes, and Trust in Communication. *Human Nature*, 31(3), 296–321. <https://doi.org/10.1007/s12110-020-09376-3>
- Schouten, J. W., McAlexander, J. H., & Koenig, H. F. (2007). Transcendent customer experience and brand community. *Journal of the Academy of Marketing Science*, 35, 357–368. <https://doi.org/10.1007/s11747-007-0034-4>

- Schroeder, J., & Schroeder, M. (2018). Trusting in Machines: How Mode of Interaction Affects Willingness to Share Personal Information with Machines. *Proceedings of the 51st Hawaii International Conference on System Sciences*. <https://doi.org/10.24251/hicss.2018.061>
- Schuetz, S., & Venkatesh, V. (2020). Research perspectives: The rise of human machines: How cognitive computing systems challenge assumptions of user-system interaction. *Journal of the Association for Information Systems*, 21(2), 460–482. <https://doi.org/10.17705/1jais.00608>
- Schuetzler, R. M., Grimes, G. M., & Giboney, J. S. (2019). The effect of conversational agent skill on user behavior during deception. *Computers in Human Behavior*. <https://doi.org/10.1016/j.chb.2019.03.033>
- Schuetzler, R. M., Grimes, M., Scott Giboney, J., Buckman, J. J., Scott, J., Grimes, G. M., Giboney, J. S., & Buckman, J. J. (2014). Facilitating natural conversational agent interactions: Lessons from a deception experiment. *35th International Conference on Information Systems "Building a Better World Through Information Systems", ICIS 2014*. <https://digitalcommons.unomaha.edu/isqafacproc/16>
- Schwartz, E. H. (2022). *Amazon Brings Alexa In-Skill Purchasing to Canada*. Voicebot.Ai. <https://voicebot.ai/2022/04/13/amazon-brings-alexa-in-skill-purchasing-to-canada/>
- Schwarz, N. (2004). Metacognitive experiences in consumer judgment and decision making. *Journal of Consumer Psychology*, 14(4), 332–348. https://doi.org/10.1207/s15327663jcp1404_2
- Schwarz, N., Jalbert, M., Noah, T., & Zhang, L. (2021). Metacognitive experiences as information: Processing fluency in consumer judgment and decision making. *Consumer Psychology Review*, 4(1), 4–25. <https://doi.org/10.1002/arcp.1067>
- Sciuto, A., Saini, A., Forlizzi, J., & Hong, J. I. (2018). “Hey Alexa, What’s Up?”: Studies of In-Home Conversational Agent Usage. *DIS 2018, June 9–13, 2018, Hong Kong*, 857–868. http://www.deutschlandfunkkultur.de/zur-zukunft-der-bibliotheken-eine-datei-ersetzt-kein-buch.1270.de.html?dram:article_id=409730
- Seaborn, K., Miyake, N. P., Pennefather, P., & Otake-Matsuura, M. (2021). Voice in Human–Agent Interaction. *ACM Computing Surveys*, 54(4), 1–43. <https://doi.org/10.1145/3386867>

- Seeber, I., Bittner, E., Briggs, R. O., de Vreede, T., de Vreede, G. J., Elkins, A., Maier, R., Merz, A. B., Oeste-Reiß, S., Randrup, N., Schwabe, G., & Söllner, M. (2020). Machines as teammates: A research agenda on AI in team collaboration. *Information and Management*, 57(2), 103174. <https://doi.org/10.1016/j.im.2019.103174>
- Seeger, A. M., Pfeiffer, J., & Heinzl, A. (2021). Texting with humanlike conversational agents: Designing for anthropomorphism. *Journal of the Association for Information Systems*, 22(4), 931–967. <https://doi.org/10.17705/1jais.00685>
- Seeger, A.-M., & Heinzl, A. (2021). Chatbots often Fail! Can Anthropomorphic Design Mitigate Trust Loss in Conversational Agents for Customer Service. *Twenty-Ninth European Conference on Information Systems (ECIS)*, 1–19.
- Sezgin, E., Huang, Y., Ramtekkar, U., & Lin, S. (2020). Readiness for voice assistants to support healthcare delivery during a health crisis and pandemic. *Npj Digital Medicine*, 3(1). <https://doi.org/10.1038/s41746-020-00332-0>
- Shevchuk, N., Degirmenci, K., & Oinas-Kukkonen, H. (2019). Adoption of Gamified Persuasive Systems to Encourage Sustainable Behaviors: Interplay between Perceived Persuasiveness and Cognitive Absorption Completed Research Paper. *Fortieth International Conference on Information Systems*.
- Singh, A., Ramasubramanian, K., & Shivam, S. (2019). Processes in the Banking and Insurance Industries. In *Building an Enterprise Chatbot* (1st ed.). Apress.
- Singh, J., Nambisan, S., Bridge, R. G., & Brock, J. K. U. (2020). One-Voice Strategy for Customer Engagement. *Journal of Service Research*, 24(1), 42–65. <https://doi.org/10.1177/1094670520910267>
- Singh, S., & Thakur, H. K. (2020). Survey of Various AI Chatbots Based on Technology Used. *ICRITO 2020 - IEEE 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)*, 1074–1079. <https://doi.org/10.1109/ICRITO48877.2020.9197943>
- Smith, S. (2018a). *CHATBOTS TO DELIVER \$11BN IN ANNUAL COST SAVINGS FOR RETAIL, BANKING & HEALTHCARE SECTORS BY 202*. Juniper Research. <https://www.juniperresearch.com/press/chatbots-to-deliver-11bn-cost-savings-2023>
- Smith, S. (2018b). *Digital voice assistants in use to triple to 8 billion by 2023, Driven by Smart Home Devices*. Juniper Research. <https://www.juniperresearch.com/press/digital-voice-assistants-in-use-to-8-million-2023>

- Smith, S. (2019). *Chatbot Interactions in Retail to Reach 22 Billion By 2023, As Ai Offers Compelling New Engagement Solutions*. Juniper Research. <https://www.juniperresearch.com/press/press-releases/chatbot-interactions-retail-reach-22-billion-2023>
- Söllner, M., Benbasat, I., Gefen, D., Leimeister, J. M., & Pavlou, P. A. (2016). Trust. In A. Bush & A. Rai (Eds.), *MIS Quaterly Research Curations*. <http://misq.org/research-curations>
- Söllner, M., Hoffmann, A., Hoffmann, H., Wacker, A., & Leimeister, J. M. (2012). Understanding the Formation of Trust in IT Artifacts. *Proceedings of the 33rd International Conference on Information Systems*. https://doi.org/10.1007/978-3-319-05044-7__3
- Söllner, M., Hoffmann, A., & Leimeister, J. M. (2016). Why different trust relationships matter for information systems users. *European Journal of Information Systems*, 25(3), 274–287. <https://doi.org/10.1057/ejis.2015.17>
- Söllner, M., & Pavlou, P. (2016). A Longitudinal Perspective on Trust in IT Artifacts. *ECIS 2016 Proceedings, 2008*.
- Son, Y., & Oh, W. (2018). “Alexa, buy me a movie!”: How AI speakers reshape digital content consumption and preference. *39th International Conference on Information Systems*.
- Song, H., & Schwarz, N. (2008). If it’s hard to read, it’s hard to do: Processing fluency affects effort prediction and motivation. *Psychological Science*, 19(10), 986–988. <https://doi.org/10.1111/j.1467-9280.2008.02189.x>
- Song, H., & Schwarz, N. (2009). If It’s Difficult to Pronounce, It Must Be Risky. *Psychological Science*, 20(2), 135–138. <https://doi.org/10.1111/j.1467-9280.2009.02267.x>
- Song, S., Baba, J., Nakanishi, J., Yoshikawa, Y., & Ishiguro, H. (2020). Mind the voice!: Effect of robot voice pitch, robot voice gender, and user gender on user perception of teleoperated robots. *Conference on Human Factors in Computing Systems - Proceedings*. <https://doi.org/10.1145/3334480.3382988>
- Srivastava, S. C., & Chandra, S. (2018). Social presence in virtual world collaboration: An uncertainty reduction perspective using a mixed methods approach1. *MIS Quarterly: Management Information Systems*. <https://doi.org/10.25300/MISQ/2018/11914>

- Steenkamp, J.-B. E. M., & Baumgartner, H. (1992). The Role of Optimum Stimulation Level in Exploratory Consumer Behavior. *Journal of Consumer Research*, 19(3), 434–448. <https://doi.org/10.1086/209313>
- Steuer, J. (1992). Defining Virtual Reality: Dimensions Determining Telepresence. *Journal of Communication*, 42(4), 73–93. <https://doi.org/10.1111/j.1460-2466.1992.tb00812.x>
- Strait, M., Vujovic, L., Floerke, V., Scheutz, M., & Urry, H. (2015). Too Much Humanness for Human-Robot Interaction. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*, 3593–3602. <https://doi.org/10.1145/2702123.2702415>
- Strand, E. A. (2000). *Gender stereotype effects in speech processing* [Ph.D. Thesis]. The Ohio State University.
- Stumpf, S., Rajaram, V., Li, L., Wong, W. K., Burnett, M., Dietterich, T., Sullivan, E., & Herlocker, J. (2009). Interacting meaningfully with machine learning systems: Three experiments. *International Journal of Human Computer Studies*, 67(8), 639–662. <https://doi.org/10.1016/j.ijhcs.2009.03.004>
- Suddaby, R. (2010). Editor's Comments : Construct Clarity in Theories of. *The Academy of Management Review*, 35(3), 346–357.
- Sundar, S. S. (2008). *The MAIN model: A heuristic approach to understanding technology effects on credibility* (pp. 73–100). MacArthur Foundation Digital Media and Learning Initiative.
- Sung, B., Vanman, E. J., & Hartley, N. (2022). Revisiting (dis)fluency: Metacognitive difficulty as a novelty cue that evokes feeling-of-interest. *Psychology & Marketing*, December 2020. <https://doi.org/10.1002/mar.21664>
- Sutton, S. J. (2020). Gender Ambiguous, not Genderless: Designing Gender in Voice User Interfaces (VUIs) with Sensitivity. *Proceedings of the 2nd Conference on Autonomous Conversational User Interfaces*. <https://doi.org/10.1145/3405755.3406123>
- Sutton, S. J., Foulkes, P., Kirk, D., & Lawson, S. (2019). Voice as a Design Material: Sociophonetic Inspired Design Strategies in Human-Computer Interaction. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/3290605.3300833>

- Sweezy, M. (2019). *Key Chatbot Statistics to Know in 2019*. Salesforce. <https://www.salesforce.com/blog/chatbot-statistics/>
- Tay, B., Jung, Y., & Park, T. (2014). When stereotypes meet robots: The double-edge sword of robot gender and personality in human-robot interaction. *Computers in Human Behavior*, 38, 75–84. <https://doi.org/10.1016/j.chb.2014.05.014>
- Thiebes, S., Lins, S., & Sunyaev, A. (2021). Trustworthy artificial intelligence. *Electronic Markets*, 31(2), 447–464. <https://doi.org/10.1007/s12525-020-00441-4>
- Thomaz, F., Salge, C., Karahanna, E., & Hulland, J. (2020). Learning from the Dark Web: leveraging conversational agents in the era of hyper-privacy to enhance marketing. *Journal of the Academy of Marketing Science*, 48(1), 43–63. <https://doi.org/10.1007/s11747-019-00704-3>
- Thompson, D. v., & Ince, E. C. (2013). When disfluency signals competence: The effect of processing difficulty on perceptions of service agents. *Journal of Marketing Research*, 50(2), 228–240. <https://doi.org/10.1509/jmr.11.0340>
- Tian, L., Moore, J. D., & Lai, C. (2017). Recognizing emotions in spoken dialogue with acoustic and lexical cues. *ISIAA 2017 - Proceedings of the 1st ACM SIGCHI International Workshop on Investigating Social Interactions with Artificial Agents, Co-Located with ICMI 2017*, 45–46. <https://doi.org/10.1145/3139491.3139497>
- Tolmeijer, S., Zierau, N., Janson, A., Wahdatehagh, J. S., Leimeister, J. M. M., & Bernstein, A. (2021). Female by Default? - Exploring the Effect of Voice Assistant Gender and Pitch on Trait and Trust Attribution. *Conference on Human Factors in Computing Systems - Proceedings*. <https://doi.org/10.1145/3411763.3451623>
- Torraco, R. J. (2005). Writing Integrative Literature Reviews: Guidelines and Examples. *Human Resource Development Review*, 4(3), 356–367. <https://doi.org/10.1177/1534484305278283>
- Tractica. (2016). *Worldwide Usage of Virtual Digital Assistant*. <https://De.Statista.Com/Statistik/Daten/Studie/620321/Umfrage/Nutzung-von-Virtuellen-Digitalen-Assistenten-Weltweit/>.
- Trevino, L. K., & Webster, J. (1992). Flow in Computer-Mediated Communication. *Communication Research*, 19(5), 539–573.
- Tricks, H. (2021). *The return of one-to-one commerce*. The Economist. <https://www.economist.com/special-report/2021/03/11/the-return-of-one-to-one-commerce>

- Tsay, C. H. H., Kofinas, A. K., Trivedi, S. K., & Yang, Y. (2020). Overcoming the novelty effect in online gamified learning systems: An empirical evaluation of student engagement and performance. *Journal of Computer Assisted Learning*, 36(2), 128–146. <https://doi.org/10.1111/jcal.12385>
- Tsiros, M., & Heilman, C. M. (2005). The effect of expiration dates and perceived risk on purchasing behavior in grocery store perishable categories. *Journal of Marketing*, 69(2), 114–129. <https://doi.org/10.1509/jmkg.69.2.114.60762>
- Turing, A. M. (1950). Computing Machinery and Intelligence. *MIND A Quaterly Review of Psychology and Philosophy*, 59(236), 433–460.
- UNESCO. (2019). *I'd blush if I could*. <https://en.unesco.org/news/first-unesco-recommendations-combat-gender-bias-applications-using-artificial-intelligence>
- Unger, L. S., & Kernan, J. B. (1983). On the Meaning of Leisure: An Investigation of Some Determinants of the Subjective Experience. *Journal of Consumer Research*, 9(4), 381–392. <https://doi.org/10.1086/208932>
- Vaccaro, K., Agarwalla, T., Shivakumar, S., & Kumar, R. (2018). Designing the Future of Personal Fashion. *Proceedings of the 2018 CHI Conference Human Factors in Computing Systems*, 1–11. <https://doi.org/10.1145/3173574.3174201>
- van Doorn, J., Mende, M., Noble, S. M., Hulland, J., Ostrom, A. L., Grewal, D., & Petersen, J. A. (2017). Domo Arigato Mr. Roboto: Emergence of Automated Social Presence in Organizational Frontlines and Customers' Service Experiences. *Journal of Service Research*, 20(1), 43–58. <https://doi.org/10.1177/1094670516679272>
- van Es, I., Heylen, D., van Dijk, B., & Nijholt, A. (2002). Gaze behavior of talking faces makes a difference. *Conference on Human Factors in Computing Systems - Proceedings*, 734–735. <https://doi.org/10.1145/506558.506571>
- van Kerrebroeck, H., Brengman, M., & Willems, K. (2017). When brands come to life: experimental research on the vividness effect of Virtual Reality in transformational marketing communications. *Virtual Reality*, 21(4), 177–191. <https://doi.org/10.1007/s10055-017-0306-3>
- van Vugt, H. C., Bailenson, J. N., Hoorn, J. F., & Konijn, E. A. (2010). Effects of facial similarity on user responses to embodied agents. *ACM Transactions on Computer-Human Interaction*, 17(2), 1–27. <https://doi.org/10.1145/1746259.1746261>

- van Zeeland, H., & Schmitt, N. (2013). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics*, 34(4), 457–479. <https://doi.org/10.1093/applin/ams074>
- Vernuccio, M., Patrizi, M., & Pastore, A. (2020). Brand Anthropomorphism and Brand Voice: The Role of the Name-Brand Voice Assistant. In *Advances in Digital Marketing and eCommerce. Springer Proceedings in Business and Economics*. (pp. 31–39). Springer.
- Viswanathan, M., & Childers, T. L. (1996). Processing of Numerical and Verbal Product Information. *Journal of Consumer Psychology*, 5(4), 359–385. https://doi.org/10.1207/s15327663jcp0504_03
- vom Brocke, J., Simons, A., Niehavens, B., Reimer, K., Plattfaut, R., & Cleven, A. (2009). Reconstructing the Giant: On the Importance of Rigour in Docmeunting the Literature Search Process. *17th European Conference on Information Systems*, 2206–2217. <https://doi.org/10.1108/09600031211269721>
- vom Brocke, J., Simons, A., Riemer, K., Niehaves, B. B., Plattfaut, R., & Cleven, A. (2015). Standing on the shoulders of giants: Challenges and recommendations of literature search in information systems research. *Communications of the Association for Information Systems*, 37, 205–224. <https://doi.org/10.17705/1CAIS.03709>
- Vtyurina, A., Savenkov, D., Agichtein, E., & Clarke, C. L. A. (2017). Exploring Conversational Search With Humans, Assistants, and Wizards. *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 2187–2193. <http://dx.doi.org/10.1145/3027063.3053175>
- Wallace, R. S. (2009). The anatomy of a.I.I.C.e. In R. Epstein, G. Roberts, & G. Beber (Eds.), *Parsing the turing test: philosophical and methodological issues in the quest for the thinking computer* (pp. 181–210). Springer.
- Walter, M., Hildebrand, C., Häubl, G., & Herrmann, A. (2020). Mixing It Up: Unsystematic Product Arrangements Promote the Choice of Unfamiliar Products. *Journal of Marketing Research*, 57(3), 509–526. <https://doi.org/10.1177/0022243720901520>
- Walther, J. B. (2005). Let Me Count the Ways: The Interchange of Verbal and Nonverbal Cues in Computer- Mediated and Face-to-Face Affinity. *Journal of Language and Social Psychology*, 24(1), 36–65. <https://doi.org/10.1177/0261927X04273036>
- Walton, D. (2014). *Abductive Reasoning*. University of Alabama Press.

- Wambsganss, T., Höch, A., Zierau, N., & Söllner, M. (2021). Ethical Design of Conversational Agents: Towards Principles for a Value-Sensitive Design. *16th International Conference on Wirtschaftsinformatik*.
- Wambsganss, T., Winkler, R., Schmid, P., & Söllner, M. (2020). Unleashing the Potential of Conversational Agents for Course Evaluations: Empirical Insights from a Comparison with Web Surveys. *Twenty-Eighth European Conference on Information Systems (ECIS 2020)*, 1–18.
- Wang, W., & Benbasat, I. (2005). Trust in and Adoption of Online Recommendation. *Journal of the Association for Information Systems*, 6(3), 72–101.
- Wang, W., & Benbasat, I. (2007). Recommendation agents for electronic commerce: Effects of explanation facilities on trusting beliefs. In *Journal of Management Information Systems* (Vol. 23, Issue 4). <https://doi.org/10.2753/MIS0742-1222230410>
- Waytz, A., Heafner, J., & Epley, N. (2014). The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology*, 52, 113–117. <https://doi.org/10.1016/j.jesp.2014.01.005>
- Webster, J., Trevino, L. K., & Ryan, L. (1993). The dimensionality and correlates of flow in human-computer interactions. *Computers in Human Behavior*, 9(4), 411–426. [https://doi.org/10.1016/0747-5632\(93\)90032-N](https://doi.org/10.1016/0747-5632(93)90032-N)
- Webster, J., & Watson, R. T. (2002). Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MIS Quarterly*, 26(2), xiii–xxiii. <https://doi.org/10.2307/4132319>
- Weizenbaum, J. (1966). ELIZA—A Computer Program For the Study of Natural Language Communication Between Man And Machine. *Communications of the ACM*, 9(1), 36–45. <https://doi.org/10.1145/357980.357991>
- Westerman, D., Cross, A. C., & Lindmark, P. G. (2019). I Believe in a Thing Called Bot: Perceptions of the Humanness of “Chatbots.” *Communication Studies*, 70(3), 295–312. <https://doi.org/10.1080/10510974.2018.1557233>
- Whang, C., & Im, H. (2021). “I Like Your Suggestion!” the role of humanlikeness and parasocial relationship on the website versus voice shopper’s perception of recommendations. *Psychology and Marketing*, 38(4), 581–595. <https://doi.org/10.1002/mar.21437>

- Wiemann, J. M., & Knapp, M. L. (1975). Turn-taking in Conversations. *Journal of Communication*, 25, 75–92. <https://doi.org/10.1111/j.1460-2466.1975.tb00582.x>
- Winkielman, P., Schwarz, N., & Reber, R. (2003). The Hedonic Marking of Processing Fluency: Implications for Evaluative Judgment. In J. Musch & K. C. Klauer (Eds.), *The Psychology of Evaluation* (pp. 205–234). Psychology Press.
- Winkler, R., Elshan, E., Söllner, M., & Leimeister, J. M. (2020). SPAM - A process model for developing smart personal assistants. *Proceedings of the Annual Hawaii International Conference on System Sciences, 2020-Janua*, 261–270. <https://doi.org/10.24251/hicss.2020.033>
- Winkler, R., Neuweiler, M. L., Leimeister, J. M., Söllner, M., & Rossini, F. C. (2019). Alexa, can you help us solve this problem? How conversations with smart personal assistant tutors increase task group outcomes. *Conference on Human Factors in Computing Systems - Proceedings*. <https://doi.org/10.1145/3290607.3313090>
- Winkler, R., & Söllner, M. (2018). Unleashing the Potential of Chatbots in Education: A State-Of-The-Art Analysis. *Academy of Management Proceedings*. <https://doi.org/10.5465/ambpp.2018.15903abstract>
- Winograd, T. (1972). Understanding Natural Language. *Cognitive Psychology*, 3, 1–191. <https://doi.org/10.3233/978-1-60750-049-0-551>
- Wirtz, J., Patterson, P. G., Kunz, W. H., Gruber, T., Lu, V. N., Paluch, S., & Martins, A. (2018). Brave new world: service robots in the frontline. *Journal of Service Management*, 29(5), 907–931. <https://doi.org/10.1108/JOSM-04-2018-0119>
- Wong, A., Xu, A., & Dudek, G. (2019). Investigating trust factors in human-robot shared control: Implicit gender bias around robot voice. *Proceedings - 2019 16th Conference on Computer and Robot Vision, CRV 2019*, 195–200. <https://doi.org/10.1109/CRV.2019.00034>
- Wong-Villacres, M., Evans, H., Schechter, D., DiSalvo, B., & Kumar, N. (2019). Con-sejero automatico: Chatbots for Supporting Latino Parents' Educational Engagement Marisol. *Proceedings of the Tenth International Conference on Information and Communication Technologies and Development*. <https://doi.org/10.1145/3287098.3287149>
- Wunderlich, N., & Paluch, S. (2017). A Nice and Friendly Chat With a Bot: User Perceptions of AI-based Service Agents. *Proceedings of the 28th International Conference on Information Systems*.

- Xiao, L., & Kumar, V. (2021). Robotics for Customer Service: A Useful Complement or an Ultimate Substitute? *Journal of Service Research*, 24(1), 9–29. <https://doi.org/10.1177/1094670519878881>
- Xu, A., Liu, Z., Guo, Y., Sinha, V., & Akkiraju, R. (2017). A new chatbot for customer service on social media. *Conference on Human Factors in Computing Systems - Proceedings, 2017-May*, 3506–3510. <https://doi.org/10.1145/3025453.3025496>
- Yu, Q., Nguyen, T., Prakkamakul, S., & Salehi, N. (2019). “I Almost Fell in Love with a Machine”: Speaking with Computers Affects Self-disclosure. *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–6. <https://doi.org/10.1145/3290607.3312918>
- Zack, W. (2020). *Robot lawyer startup DoNotPay now lets you file FOIA requests*. TechCrunch. https://techcrunch.com/2020/12/10/robot-lawyer-startup-donotpay-now-lets-you-file-foia-requests/?guccounter=1&guce_referrer=aHR0cHM6Ly9uZXdzLmdvb2dsZS5jb20v&guce_referrer_sig=AQAAA-BEE5P9GkxYx1A67XrMedJctPTaeRDcmB-BZlldh2t3KXhUYdzEBF5hoO8xpSXc5uQhQaGf-_UI2U
- Zanjani, S. H. A., Milne, G. R., & Miller, E. G. (2016). Procrastinators’ online experience and purchase behavior. *Journal of the Academy of Marketing Science*, 44(5), 568–585. <https://doi.org/10.1007/s11747-015-0458-1>
- Zellner, A. (1962). An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias. *Journal of the American Statistical Association*, 57(298), 348–368.
- Zhang, S. (2015). *The “Harvard Sentences” Secretly Shaped the Development of Audio Tech*. GIZMODO. <https://gizmodo.com/the-harvard-sentences-secretly-shaped-the-development-1689793568>
- Zierau, N., Elshan, E., Visini, C., & Janson, A. (2020). A Review of the Empirical Literature on Conversational Agents and Future Research Directions. *International Conference on Information Systems (ICIS)*.
- Zierau, N., Engel, C., Söllner, M., & Leimeister, J. M. (2020). Trust in Smart Personal Assistants: A Systematic Literature Review and Development of a Research Agenda. *WI2020 Zentrale Tracks*, 99–114. https://doi.org/10.30844/wi_2020_a7-zierau

- Zierau, N., Hildebrand, C., Bergner, A., Busquet, F., Schmitt, A., & Marco Leimeister, J. (2022). Voice bots on the frontline: Voice-based interfaces enhance flow-like consumer experiences & boost service outcomes. *Journal of the Academy of Marketing Science*, 0123456789. <https://doi.org/10.1007/s11747-022-00868-5>
- Zillermann, D., & Brosius, H.-B. (2012). *Exemplification in Communication*. Routledge.
- Zomerdijk, L. G., & Voss, C. A. (2010). Service design for experience-centric services. *Journal of Service Research*, 13(1), 67–82. <https://doi.org/10.1177/1094670509351960>

Curriculum Vitae

Name	Naim Simon Abbas Zierau
Date of birth	03.05.1989
Place of birth	Bremen, Germany
Nationality	German

Education

2018-2022	University of St. Gallen (HSG), St. Gallen, Switzerland Ph.D. Program in Management (Business Innovation)
2015-2018	Karlsruhe Institute of Technology, Karlsruhe, Germany M.Sc. Industrial Engineering and Management
2009-2013	Technical Univsity Darmstadt, Darmstadt, Germany B.Sc. Business Engineering
2008	Leibniz Gymnasium, Bad Schwartau, Germany Abitur

Professional Experience

2018-2022	IWI-HSG, St. Gallen, Switzerland Research Assistant
2016-2018	Karlsruhe Service Research Institute Design Thinking Coach
2016	Mensch GmbH, Oxford, United Kingdom Management Consulting Intern
2014-2015	Ruhi Institute, Sofia, Bulgaria Program Coordination
2013-2014	HeidelbergCement AG, Heidelberg, Germany Project Managemen Intern