# Designing Relatable AI:

## A Theory of Mind Perspective of How Conversational AI Shapes Technology-Mediated Consumer Behavior

D I S S E R T A T I O N
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

**Anouk Samantha Bergner**

from

Biel (Bern)

Approved on the application of

**Prof. Dr. Christian Hildebrand**

and

**Prof. Dr. Damian Borth**

Dissertation no. 5163

Difo-Druck GmbH, Untersiemau 2022

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

St.Gallen, October 29, 2021

The President:

Prof. Dr. Bernhard Ehrenzeller

*To Mila and Vince*

# Acknowledgements

The three years leading up to this dissertation have been an intense journey of intellectual discoveries and personal growth. They would not have been possible without the support from and contribution of many amazing people along the way. First, I would like to express my sincere gratitude to my advisor, Prof. Dr. Christian Hildebrand, for his inspiring guidance and unwavering trust in me. I would also like to thank my co-advisor, Prof. Dr. Damian Borth for his helpful comments on my research proposal and his guidance on our deep learning classifier. Finally, I would also like to thank Prof. Dr. Danny Oppenheimer, for encouraging me to go back to academia and pursue my PhD and his continued feedback throughout this incredible experience.

Next, I would like to thank my amazing co-authors Prof. Dr. Gerald Häubl and Dr. Jochen Hartmann for the outstanding collaboration on our projects, and the guidance they have provided throughout the research and writing of our joint papers. I have learnt so much from them, both personally and professionally. I would also like to thank the entire TechX team for their continuous feedback during numerous lab meetings, which have shaped and influenced each one of my projects. Will, Fran, Fotis, Meike, Anna, and Ilaria, you have been the most amazing team members and colleagues, provided me with technical and moral support whenever needed, and I have tremendously enjoyed our intense conversations and jumping into freezing mountain lakes with you. Finally, I would also like to thank Doris and Claudio for their important work behind the scenes.

Last but certainly not least, I consider myself beyond lucky to have the most amazing family and incredible friends around the world that I can always count on for encouragement and support. Thank you especially to my parents, Rolf and Fränzi and my sister Franca for always being there for me, to my husband Greg for supporting me every day and pushing me to pursue what I love, and to our wonderful children Mila and Vince, for being the sunshine of my life. I hope they will forgive me for the long hours spent glued to my laptop and I will inspire them to reach high and never give up.

Geneva, January 13, 2022 Anouk S. Bergner

# Abstract

As technology is evolving to appear progressively more human-like, next generation conversational artificial intelligence (AI) is profoundly impacting human-technology relationships. Conversational interfaces (CIs) effectively mimic human conversation in either written (as with chatbots) or spoken form (as with digital voice assistants), taking over a broad range of tasks from acting as personal assistants, to selling products to consumers, or recommending investment strategies in the financial industry. Despite this fascinating and broad range of applications, one unifying element across the evolving conversational AI landscape is the ability to interact with consumers in what was previously considered the hallmark of humanity: language and dialogue. This raises foundational questions of how consumers perceive progressively more humanized AI, as well as the behavioral consequences for consumers. This dissertation develops a novel theoretical framework of human-technology relationships in the context of conversational AI, building on and integrating fundamental research on the theory of mind in psychology, foundational principles of human dialogue in linguistics, and emerging literature on technology-mediated consumer behavior. The framework and the underlying hypotheses are developed and tested in four independent paper projects, across a broad range of methodological approaches (deep contextual language models and experimental research designs), study settings (both in the field and the lab), and participant samples (from financial investors to online shoppers). Together, these papers demonstrate how interacting with conversational AI fundamentally shapes consumer decision-making and assess the underlying psychological mechanisms and moderating factors influencing the formation of intimacy and trust with such interfaces (and the brands or firms they represent). Furthermore, they provide novel methodological avenues, compared to traditional consumer research, to unobtrusively identify variations in the attribution of a "mind" to these interfaces. The overarching objective of this dissertation is to contribute to an emerging and interdisciplinary field of research on how humans relate to and are influenced by increasingly humanized AI in a digitally transforming society.

# Zusammenfassung

Die rasche Entwicklung von Technologien wie künstlicher Intelligenz (KI) auf immer menschlichere Art zu kommunizieren, hat tiefgreifenden Einfluss auf die Beziehung zwischen Mensch und Technologie. Dialog-basierte KI in Form von Conversational Interfaces (CIs) ahmen menschliche Konversationen in schriftlicher (wie bei Chatbots) oder gesprochener Form (wie bei digitalen Sprachassistenten) nach und übernehmen ein breites Spektrum an Aufgaben: von der Funktion als persönlicher Assistent über den Verkauf von Produkten bis hin zur Empfehlung von Anlagestrategien in der Finanzbranche. Obwohl der Anwendungsbereich dialog-basierter KI sehr breit ist, gibt es ein übergreifendes Element, welches diese Anwendungen vereinigt: ihre nahezu menschliche Fähigkeit, mit Konsumenten in natürlicher Sprache zu interagieren. Dies wirft grundlegende Fragen darüber auf, wie diese zunehmend menschlichere KI wahrgenommen wird und welche Konsequenzen dies für das Verhalten der Konsumenten hat. Die vorliegende Dissertation entwickelt ein neues theoretisches Fundament zur Erforschung der Beziehung zwischen Mensch und KI im Kontext der CI basierend auf Literatur zur psychologischen "Theory of Mind", Grundprinzipien menschlichen Dialogs in der Linguistik und der Forschung des technologie-getriebenen Konsumentenverhalten. Das theoretische Fundament und die daraus abgeleiteten Hypothesen werden in vier unabhängigen Forschungsprojekten entwickelt und mithilfe mehrerer methodischen Ansätze (von Deep Learning Sprachmodellen bis hin zu experimentellen Forschungsdesigns), Studiensettings (sowohl Feld- als auch Laborstudien) und Stichproben (von Finanzinvestoren bis hin zu Online-Käufern) getestet. Zusammen zeigen diese Arbeiten wie die Interaktion mit dialog-basierter KI die Entscheidungsfindung von Konsumenten grundlegend beeinflusst. Sie erforschen die zugrundeliegenden psychologischen Mechanismen und moderierende Faktoren, welche die Bildung von Intimität und Vertrauen mit dialog-basierter KI (und den Marken oder Unternehmen die sie repräsentieren) beeinflussen. Sie bieten im Vergleich zur traditionellen Konsumentenforschung neuartige Methoden, um die Zuschreibung eines Menschlichen „Verstandes" zu dialog-basierter KI valide zu identifizieren. Das übergreifende Ziel dieser Dissertation ist es, einen Beitrag zu einem sich neu entwickelnden und interdisziplinären Forschungsfeld zu leisten, das sich mit der Frage beschäftigt, wie Menschen in einer sich digital transformierenden Gesellschaft mit zunehmend menschlicher KI umgehen und von dieser beeinflusst werden.

# Table of Contents

# List of Figures

# List of Tables

# 1. DISSERTATION OVERVIEW

## 1.1 Motivation and State of Research

### 1.1.1 The Rise of Conversational AI

Conversational AI represents a paradigm shift in the human-technology relationship. Advances in natural language processing (NLP) and machine learning allow a progressively more natural interaction with technology, approaching Turing's (1950) definition of machine intelligence as the inability to discern an interaction with a machine from one with another human being. The act of being able to understand and dynamically respond to user input using natural language fundamentally changes how humans interact with and relate to technology. Rather than being acted upon, technology becomes an active contributor to an interaction through the means of communication, an ability previously considered exclusively human.

Interaction with conversational AI is entering all aspects of consumers' lives from automating service operations with messenger-based customer support chatbots to executing daily tasks with voice-based virtual assistants. Recent industry reports predict that the conversational AI market continues to grow at an annual rate of 30.2%, expected to reach $17.5 billion dollars by 2024 (Deloitte 2019). Relatedly, the adoption of conversational interfaces is expected to see a marked increase in the next two to five years thanks to both, the significant evolution in the technology underlying conversational AI, as well as an increase in the adoption of digital technologies more broadly (Gartner 2021; Columbus 2020). This rapid evolution of the conversational AI landscape is accentuating its growing influence on consumers' interactions with technology and consumption experiences, fundamentally changing how consumers search for information, shop for products, and interact with firms on a daily basis.

The rise in conversational AI is reflected in a growing research interest related to the design (Landis 2014; Lokman and Zain 2010; Serban et al. 2017) and use (Araujo 2018; Go and Sundar 2019; Lee, Lee, and Sah 2019; Zarouali et al. 2018) of such interfaces, and how they may shape consumer experiences in marketing (Araujo 2018; Van den Broeck, Zarouali, and Poels 2019; Luo et al. 2019; Mende et al. 2019), finance (Hohenberger, Lee, and Coughlin 2019; Tedesco 2015), or health (Bellur and Sundar 2017; Comendador et al. 2015;

Miner et al. 2016; Oh et al. 2017). However, academic research has been sparse on how the unique property of interacting with these interfaces through natural dialogue may impact consumer behavior and the underlying psychological mechanisms shaping consumers' perception of and relationship with these interfaces. This dissertation aims at contributing to and extending this emerging and interdisciplinary field of research by investigating the underlying mechanisms and key boundary conditions of how humans relate to and are influenced by conversational AI and downstream consequences on technology-mediated consumer behavior. Given the unique aspect of human-technology interaction through natural language, this dissertation develops a novel theoretical framework for these emerging human-technology relationships in the context of conversational AI, building on and integrating fundamental research in interpersonal psychology and theory of mind (Airenti 2018; Epley, Waytz, and Cacioppo 2007; Gray, Gray, and Wegner 2007; Waytz, Cacioppo, and Epley 2010; Yang, Aggarwal, and McGill 2020), computational linguistics (Jurafsky and Martin 2017; Levinson 2016), and the emerging literature on consumer-technology relationships and technology-mediated choices (Hoffman and Novak 2018; Melumad et al. 2020; Novak and Hoffman 2019).

### 1.1.2 Definition of Conversational Interfaces

In this dissertation, AI-based conversational interfaces are defined as machine-based, dynamic dialogue systems that mimic human-to-human conversation through a sequential process of turn-taking and grounding between the user and the interface. Conversational interfaces can be either text-based or voice-based (Sheehan, Jin, and Gottlieb 2020) and the current theorizing is agnostic to differences in input modalities (written vs. spoken), as the fundamental structural and semantic properties of conversation generalize across these modalities. Thus, a conversational interface's ability to interact in natural language with a human user and dynamically adapt to his or her input distinguishes these interfaces from other interfaces, where input is provided by clicking on or touching icons or writing syntax-specific code.

### 1.1.3 Why Designing Relatable AI Matters

Seminal research on human-computer interaction has long demonstrated that humans apply social heuristics even to computers that possess only subtle human cues, such as a gender or a human voice (Moon 2000; Nass and Moon 2000; Nass, Steuer, and Tauber 1994).

With the advancement in NLP and the proliferation of CIs, machines are entering a territory that was previously exclusive to humans, mimicking human dialogue. An active conversational interaction radically changes how humans relate to technology, because the technological medium is no longer just a channel, but rather becomes a distinct source of communication and information, automatically triggering social behavior in response to the medium (Kim and Sundar 2012; Sundar and Nass 2000).

The perspective of theory of mind in this context is relevant as it relates to a fundamental aspect of social cognition, shaping how we perceive another entity and influencing how we relate to it through the means of communication (Airenti 2018). As such, the attribution of a "mind" (e.g. in the form of internal states and intentions) to technology is independent of the conscious beliefs about the technology, and automatically results from a desire to understand and relate to it during an interaction (Epley et al. 2007; Kim and Sundar 2012; Waytz, Cacioppo, et al. 2010; Yang, Aggarwal, and Mcgill 2020). This is particularly relevant in the context of AI-based CIs, allowing consumers to directly communicate with them using natural language, as it elevates the technology to one of the most fundamental human capabilities: dynamic, two-way dialogue.

This active participation in the conversational process makes technology a contributor to the consumption experience and an active partner in the decision process. However, design practitioners and brand managers are lacking knowledge on how conversational properties of AI-driven CIs may fundamentally shape these experiences and influence consumer choice. More specifically, previous research has primarily focused on the influence of identity and visual characteristics of the interface (Bertacchini, Bilotta, and Pantano 2017; Holzwarth, Janiszewski, and Neumann 2006; Jin and Sung 2010; Wang et al. 2007), rather than how conversations can essentially be "engineered" to nurture intimacy and trust with the technology and the brand or firm it represents. Furthermore, while more relatable AI can be beneficial to the consumer experience, leading to more personalized, one-to-one consumer-firm interactions, it may also be detrimental to the consumer if the firm misuses the increase in trust to its advantage, implying a potential need for consumer protection.

This dissertation proposes a novel conceptualization of conversational AI based on the unique properties of language and natural dialogue and develops several hypotheses of how these interfaces can fundamentally alter consumer perceptions and technology-mediated choice. The findings presented in this dissertation have important design implications for practitioners, public policy implications for consumer welfare, and theory implications for the emerging field of the consumer psychology of new technologies.

## 1.2 Theoretical Background

### 1.2.1 Theory of Mind

Thinking about internal mental states of another entity, especially another human, is a well-studied phenomenon in psychology (Abell, Happé, and Frith 2000; Airenti 2018; Bretherton and Beeghly 1982; Epley et al. 2007; Goldstein and Winner 2012; Premack and Woodruff 1978) and is considered a fundamental human capacity. Developing such a *theory of mind* is crucial to human interaction and serves the function of making inferences about another entity's behavior. As virtually all human behavior is driven by underlying mental states, humans learn early on in childhood how to make inferences about the internal states of another being, such as its intentions, knowledge, and beliefs, in order to make sense of its actions (Birch et al., 2017).

This reasoning about internal, unobservable states in other entities has been shown to extend far beyond human-to-human interaction to reasoning about our pets' behavior (Paul et al. 2014) and even the behavior of objects (Abell et al. 2000; Epley et al. 2007; Waytz, Cacioppo, et al. 2010) and tends to unfold naturally and automatically, fundamentally shaping how we perceive and relate to others (Airenti 2018; Gray et al. 2007; Waytz, Gray, et al. 2010). In contrast, denying such essential human capacities leads to a sense of dehumanization, in which entities lacking these human features are likened to machines and perceived as cold, passive, and shallow (Haslam et al., 2008). Hence, theory of mind is a fundamental aspect of social perception and the attribution of human traits and intentions (both consciously and unconsciously) is one of the primary media through which humans define their relationship with another entity.

As technology is evolving to interact and communicate in a progressively more human way, theory of mind is central to human-technology relationships because it is based on a learnt assumption that human behavior is driven by stable underlying traits and values. Epley et al. (2007) propose that our propensity to attribute human traits and mental states to other entities is influenced by both, cognitive and social factors. Specifically, the cognitive determinant is linked to the ease at which humans automatically access information about themselves and the tendency to therefore compare an entity's characteristics and behaviors to a known human schema. Social factors on the other hand, focus on our interpretation of the behavior of the entity in relationship to its environment, as well as ourselves. Humans strive to understand and predict the behavior of another entity in order to reduce uncertainty about their environment, and often achieve this by trying to take its perspective, thus, essentially leveraging theory of mind in an attempt to understand and relate to the other entity.

### 1.2.2 The Role of Language and Dialogue

A strong connection between human communication and theory of mind has long been established in developmental psychology, demonstrating the existence of important links between language acquisition and the ability to reason about another person's mental state and behavior (Astington and Baird 2005). There are two importance consequences of the strong ties between theory of mind and language, both addressed in this dissertation. First, the fact that consumers can talk *with* CIs using natural language and dialogue should increase consumers' attribution of human traits to these interfaces, the more the interface embodies fundamental principles of human dialogue (addressed in Articles I, II, and III). Indeed, research has shown that a child's participation in two-way conversations leads to his or her attribution and understanding of mental states in other humans (Bretherton, McNew, and Beeghly-Smith 1981; Dunn et al. 1991). Therefore, actively conversing with an interface using natural dialogue should fundamentally impact perceptions of the interface and increase attributions of humanness to it.

Second, children's theory of mind development is directly linked to their acquisition of specific lexical terms used to refer to mental states of the other entity (Bretherton and Beeghly 1982). This suggests that there may be subtle cues in the way in which consumers talk *about* CIs that would allow identifying and understanding the attribution of such a "mind" to CIs and gain insight into the underlying relationship with it (addressed in Article IV). Thus, as CIs are entering the realm of natural language and dialogue, they are fundamentally redefining how humans perceive and relate to technology.

### 1.2.3 Establishing Relationships

One of the key elements of establishing a relationship is the decision to trust the other person. Trust is often defined as a willingness to take the risk of relying on another party with the expectations that the other party will act in that person's best interest (Glikson and Woolley 2020; Hoff and Bashir 2015). Previous research suggests that there are two distinct types of trust: cognitive-based trust, which is based on a rational evaluation of the competence and reliability of the other party and affect-based trust, which develops as a result of the social connection, intimacy, and positive affect with the other party (Glikson and Woolley 2020; McAllister 1995). While previous research in technology has focused primarily on cognitive-based trust in the adoption of new technologies (Davis 1989) more recent work has found that emotions are an important, and often primary, determinant of trust in automation (Hoff and Bashir 2015; Lee and See 2004).

As human-technology interactions become progressively more natural, it can be more difficult to cognitively and rationally understand the technology, implying that an emotional connection and affect-based trust will become more important in the evaluation of such technologies. Furthermore, while cognitive trust by definition relies on a more objective evaluation of the other party (for example, based on its reliability, dependability, or transparency; c.f. Hoff and Bashir 2015), affect-based trust relies on a more subjective and emotional evaluation, assessing the underlying internal motivation and intentions of the other party. Given that theory of mind is the fundamental process through which humans assess internal states and behavioral intentions of another entity, it is expected to play a fundamental role in the perception of AI-based conversational interfaces and the establishment of an emotional connection with it.

### 1.2.4 Technology-Mediated Consumer Behavior

The ability of AI-based CIs to dynamically interact, recommend options and execute tasks makes them active partners in the consumption process (Hoffman and Novak 2018; Melumad et al. 2020). As such, while CIs are merely a technological medium to complete a task, the joint activity of conversation is expected to profoundly influence, not only the perception of the interface, but also resulting relationships with the brand and firms they represent, as well as downstream consumer behavior.

As greater levels of relationship closeness and trust have been shown to increasing cooperation and coordinated social interactions (Gächter, Herrmann, and Thöni 2004; Rousseau et al. 1998), it is expected that conversational AI will influence consumers' willingness to collaborate and coordinate activities, allowing CIs to contribute to, or even take-over decision-processes. Furthermore, research on persuasion suggests that greater intimacy and affective trust increases people's susceptibility to influence attempts (Aronson 1999; Cialdini and Sagarin 1994; Finch and Cialdini 1989; Fransen, Smit, and Verlegh 2015; Goldstein, Cialdini, and Griskevicius 2008; Maner et al. 2002) and elevates persuasion effectiveness more broadly (Betancourt 1990; Hexmoor, Rahimi, and Chandran 2008). Thus, not only are consumers likely to allow CIs they trust more to execute certain tasks on their behalf, but this felt increase in closeness and intimacy can potentially have important consequences on the persuasiveness of the interface and acceptance of recommendations made by it. This can be to the benefit of the consumer, as long as his/her preferences are aligned with those of the firm providing the interface. However, it may be detrimental for consumers if this increase in affective trust and intimacy is exploited to the benefit of a firm by guiding them toward choices that, while beneficial to the firm, are sub-optimal, more costly, or even entirely wrong for the consumer (Melumad et al. 2020).

Taken together, this dissertation leverages a theory of mind perspective to shed light on the perception of conversational AI and its influence on the emotional and intimate connection with, and affect-based trust placed in, such interfaces, assessing potential boundary conditions at the conversational, individual, and task level. Furthermore, the dissertation explores a broad range of resulting downstream consumer behavior ranging from more immediate behavior specific to the interaction, such as consumers' willingness to delegate tasks (Article IV) or to accept recommendations from the CI (Articles I and III), to longer term behavior that goes beyond the interaction, such as consumers' ongoing commitment to the brand or firm represented by the CI and their willingness to engage with the CI in the future (Article I, II, and III).

## 1.3 Research Strategy

### 1.3.1 Key Research Objectives

This dissertation aims at shedding light on how interacting with conversational AI fundamentally shapes consumer decision-making, assessing the underlying psychological mechanisms influencing the perception of, and relationship with, such interfaces (and the brands or firms they represent). Furthermore, the dissertation assesses a series of critical boundary conditions at the interface, individual, and task level and their influence on technology-mediated consumer behavior. Ultimately, the dissertation aspires to further our understanding of humanized AI, the important role of consumers' attribution of a "mind" toward conversational interfaces, and the downstream impact on consumer trust formation, brand relationships and consumer choice. The dissertation also leverages and combines novel methodologies (deep contextual language models) and experimental research designs using real consumer-CI interactions, to provide theoretically important and practically relevant insights. Ultimately, this dissertation aims to contribute to an emerging and interdisciplinary field of research on how humans form relationships with and are influenced by increasingly humanized AI, focusing on the context of CIs.

### 1.3.2 Overview of Projects

The dissertation consists of the accumulated research across three empirical paper projects, as well as one conceptual article. Table 1 provides an overview of the key research question, methodology, and current publication status for each of the articles, and sections 3.2.1 to 3.2.4 present a synthesis of the key findings and implications for each research project in this dissertation.

**Table 1-1:** Dissertation overview and research approach by project

| | ARTICLE I | ARTICLE II | ARTICLE III | ARTICLE IV |
|---|---|---|---|---|
| **TITLE** | Machine Talk: How Conversational Interfaces Promote Brand Intimacy and Influence Consumer Choice | Adaptive Sales Automation: Chatbots as Personalized and Scalable Sales Agents | Conversational Robo Advisors as Surrogates of Trust: Onboarding Experience, Firm Perception, and Consumer Financial Decision Making | Conferring Minds to Machines: A Deep Learning Approach to Mind Perception, Smart Object Relationships, and Task Delegation |
| **CO-AUTHORS** | Christian Hildebrand<br>Gerald Häubl | - | Christian Hildebrand* | Jochen Hartmann<br>Christian Hildebrand |
| **RESEARCH QUESTION** | How do conversation properties of AI-based conversational interfaces influence consumer-brand relationships and brand-related behavior? | How can AI-based conversational interfaces be better integrated in a customer journey to nurture customer relationships and increase sales? | Do conversational robo advisors increase trust in a financial firm and can they lead investors to accept potentially wrong portfolio recommendation? | Can subtle linguistic features reveal differences in the perception of, and relationship with, an AI-based conversational interface, and predict differences in consumer behavior? |
| **METHODOLOGY** | Experimental research design using real consumer-CI interactions | Conceptual integration of existing interdisciplinary literature and own research | Experimental research designs using real consumer-CI interactions with active financial investors | Deep contextual language model, LDA, Lasso regression, word frequency analysis, LIME algorithm, OLS regression models on unstructured customer reviews |
| **DATA** | 1 field and 4 online experiments, including 2,938 participants | Data from existing literature and 2 online experiments with 217 participants | 4 online experiments with 1,207 active financial investors | 2,592 survey participants and 29,363 scraped customer reviews |
| **PUBLICATION STATUS** | Under review (third round) at the *Journal of Consumer Research* | Published in *Marketing Review St. Gallen* | Published in the *Journal of the Academy of Marketing Science* | Ready for submission at the *Journal of Consumer Research* |
| **CONFERENCES & AWARDS** | ACR 2018, SCP 2019, ACR Special Session 2019, TPM 2019, EMAC 2020; Selected as *Best Paper Finalist* at SCP 2019 | - | ACR 2018, SCP 2019, ACR Special Session 2019 | ACR 2020, SCP 2021, EMAC 2021, AIM 2021 |

**first author*

**ARTICLE I: Machine Talk: How Conversational Interfaces Promote Brand Intimacy and Influence Consumer Choice**

***Research Objective:*** To provide a rigorous conceptualization and examination of how specific conversational properties of AI-based CIs influence the perception of the interface and affect consumer-brand relationships and brand-related behavior.

***Key Results:*** Article I identifies and critically assesses fundamental structural (turn-taking, turn-initiation) and semantic (grounding) properties of a dialogue as the key drivers of consumers' perception of a CI and resulting brand relationships. Converging evidence from one pilot and four main studies shows that these conversational properties profoundly impact the perceived humanness of conversational interfaces, which promotes more intimate consumer-brand relationships, ultimately resulting in more favorable behavioral brand outcomes (such as greater recommendation acceptance, willingness to pay a price premium, brand advocacy, and brand loyalty). Moreover, Article I reveals that these effects are enhanced in contexts requiring greater mutual understanding between the consumer and the brand.

***Implications:*** Article I highlights how our understanding of fundamental principles of human-to-human communication can be harnessed to design more relatable AI. Specifically, the article provides and rigorously tests a theory-driven framework that advances our understanding of how conversational AI influences the formation of consumer-brand relationships by leveraging fundamental principles of human communication (turn-taking, turn-initiation, and grounding). The findings have important practical implications for the effective design of AI-based CIs that brands can leverage to deliver highly scalable, one-to-one consumer-brand experiences, nurture consumer-brand relationships, and systematically influence a broad range of behavioral brand outcomes (from greater recommendation acceptance to increased brand loyalty).

**ARTICLE II: Adaptive Sales Automation: Chatbots as Personalized and Scalable Sales Agents**

***Research Objective***: To provide a conceptual framework of how CIs can be leveraged to nurture more personalized consumer-firm relationships and increase sales.

***Key Results***: Article II provides a conceptual mapping of key design features of CIs along the customer journey to propose a theory-driven design-framework aimed at humanizing consumer-CI interactions and better integrating CIs in the sales process. The article proposes clear strategies how to dynamically adapt relevant CI design features to specific brand characteristics, customer needs, and buying tasks. The conceptual model is supported with empirical evidence from two online experiments demonstrating that more personalized, adaptive CIs increase consumers' willingness to trade-up to more premium product options.

***Implications:*** Article II outlines important strategies for brand managers and marketing professionals how to design more effective CIs that can nurture closer consumer-brand relationships and adapt to specific consumer needs throughout the sales journey. The article also highlights important implications on sales automation and the effective management of CIs as an integral part of a firm's branding ecosystem and sales force, rather than treating CIs as simple cost-saving tools.

**ARTICLE III: Conversational Robo Advisors as Surrogates of Trust: Onboarding Experience, Firm Perception, and Consumer Financial Decision Making**

***Research Objective:*** To provide an industry application of CIs to the financial sector and examine how automated financial advisory, in which consumers engage in a dynamic, dialogue-based interaction with a conversational robo advisor as opposed to static, non-conversational robo advisors, can alter perceptions of affective trust, the evaluation of a financial services firm, and consumer financial decision making.

***Key Results:*** Article III reveals important consequences of interacting with conversational AI on human-technology trust formation and investor behavior in the financial sector. Evidence across four studies demonstrates that conversational robo advisors cause greater levels of affective trust and a more benevolent evaluation of a financial services firm, with potentially detrimental consequences to financial investors. Specifically, Article III shows that, as a result of the greater affective trust in the interface and benevolence attributions to the firm, investors are more prone to accept objectively wrong portfolio recommendations and select offers with larger annual management fees.

***Implications:*** Article III reveals the potential dark side of designing more relatable AI and how consumer behavior can be influenced "by design," leveraging the capacity of these interfaces to increase affective trust and leading to an implicit assumption that the firm will act in the consumer's best interest. The findings presented in Article III highlight important implications for public policy in the digital economy, suggesting a potentially greater need for regulation and consumer protection when interacting with CIs in the future.

**ARTICLE IV: Conferring Minds to Machines: A Deep Learning Approach to Mind Perception, Smart Object Relationships, and Task Delegation**

***Research Objective***: To provide a method that can automatically detect differences in the perception of and relationship with AI-based CIs and assess their influence on downstream consumer behavior.

***Key Results***: In Article IV, a state-of-the-art deep contextual language model is trained using transfer learning to unobtrusively classify the attribution of a "mind" to a CI from unsolicited, consumer-generated text. A rich set of interpretable linguistic features is exposed using manifold methods, demonstrating that the extent of mind perception provides novel insights into consumer-technology relationships and successfully predicts differences in consumers' willingness to delegate a broad set of tasks to it.

***Implications:*** Article IV provides support for theory of mind as an important determinant of consumers' perception of, and a central factor in the formation of a relationship with, conversational AI-enabled devices. Furthermore, the article proposes a novel approach, leveraging the latest technological innovations in NLP, to uncover and interpret nuanced linguistic signals of such attributions of a "mind" to CIs in unstructured text, opening up new avenues to conduct research and develop theory in marketing as well as providing a novel tool for practitioners to unobtrusively assess interaction experiences with CIs.

## 1.4 Contributions and Future Outlook

This dissertation seeks to integrate previously disconnected streams of fundamental research in psychology, linguistics, and an emerging field of technology-augmented choice, providing evidence of how different modalities (natural language, dynamic dialogue) and modifiers (virtual assistants, sales bots, and robo advisors) can alter consumption experiences and consumer decision-making (c.f. Melumad et al. 2020). Across the four articles, it becomes evident that conversational interfaces present a paradigm shift in consumer-technology relationships. Even though these interfaces act merely as a medium to complete a task (such as ordering a product, playing music, or signing up to a financial service), consumers tend to form a relationship with the medium which extends far beyond the medium itself, and toward the object, brand or firm represented by the medium, with important consequences on consumer behavior. The four articles highlight important and broad implications on how to improve consumer experiences on the one side, by making technology more relatable and understanding, nurture stronger, more trusting consumer-firm relationships, and enable a better cooperation between consumers and machines. On the other side, they also reveal a greater need for regulation and consumer protection in order to design relatable AI that is ethical and protects consumers' best interests. The present research thus contributes to an emerging field of research on the psychology of technology and technology-augmented brand relationships and consumer behavior, and opens up several interesting and unanswered questions that might inspire future work in this domain.

First, the current dissertation explicitly states that it is agnostic to the modality (written or spoken) used by the conversational interface, as the fundamental structural and semantic properties of dialogue explored in the present articles apply to both modalities. Thus, while the present dissertation looked at both text-based (Articles I, II, and III) and voice-based (Article IV) interfaces, it treated them separately, and did not investigate potential differences driven by the expression modality of these interfaces. Yet, recent research reveals that expression modality can fundamentally alter perceptions (Schroeder, Kardas, and Epley 2017) and influence preferences (Klesse, Levav, and Goukens 2015). Thus, future research may further explore potential differences in the mechanism and boundary conditions specific to the modality of text as opposed to voice-based CIs.

Second, there is potential to extend the present work on designing relatable AI to contexts beyond marketing, and leveraging the insights to overcome resistance to AI in other domains (c.f. Dietvorst, Simmons, and Massey 2018; Granulo, Fuchs, and Puntoni 2019; Longoni, Bonezzi, and Morewedge 2019). Given that some of this resistance may be driven by people's lay beliefs about algorithms as cold and distant, designing more relatable AI may help to overcome some of this resistance by increasing intimacy and affective trust in it and designing interaction experiences that are more personalized and engaging for users (initial evidence of this is discussed in the GD of Article I).

Third, there is an opportunity for future work to assess the impact of language cues and conversational properties and how they may be leveraged by CIs to express specific "personality traits". While the present dissertation focuses primarily on conversational properties that humanize an interaction and linguistic cues that reveal such perceptions of humanness in these interfaces, the language used by these interfaces may be leveraged in more nuanced ways to express specific aspects of personality, such as for example extroversion/introversion (Nass et al. 1995), challenging as opposed to nurturing intentions (Danescu-Niculescu-Mizil et al. 2012) or reflecting the extent of power of the individual (Tapus and Mataric 2008). Furthermore, these personality traits may either be designed consistent with the personality of the brand, or engineered to dynamically adapt to the consumer interacting with the interface, and are expected to have important consequences on how consumers relate to and are influenced by CIs.

Finally, thanks to their ability to acquire information about consumers either explicitly, through the conversational interaction, or unobtrusively by leveraging automated text analysis (c.f. Berger et al. 2020; Humphreys and Wang 2018), conversational interfaces provide immense opportunities to dynamically adapt to consumers during an interaction and deliver more personalized consumer experiences. Leveraging the latest contextual language models to uncover and interpret nuanced linguistic signals in unstructured text (as presented in Article IV), opens up new research avenues on how to leverage these state-of-the-art models not only to gain novel insight into consumer experiences, but to build technology that can better understand and relate to humans in the future.

## References

Abell, F., F. Happé, and U. Frith (2000), "Do Triangles Play Tricks? Attribution of Mental States to Animated Shapes in Normal and Abnormal Development," *Cognitive Development*, 15(1), 1–16.

Airenti, Gabriella (2018), "The Development of Anthropomorphism in Interaction: Intersubjectivity, Imagination, and Theory of Mind," *Frontiers in Psychology*, 9, 1–13.

Araujo, Theo (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, 183–89.

Aronson, Elliot (1999), "The Power of Self-Persuasion," *American Psychologist*, 54(11), 875–84.

Astington, Janet Wilde and Jodie A Baird (2005), *Why Language Matters for Theory of Mind*, New York, NY, US: Oxford University Press.

Bellur, Saraswathi and S Shyam Sundar (2017), "Talking Health With a Machine : How Does Message Interactivity Affect Attitudes and Cognitions ?," *Human Communication Research*, 43, 25–53.

Berger, Jonah, Ashlee Humphreys, Stephan Ludwig, Wendy W. Moe, Oded Netzer, and David A. Schweidel (2020), "Uniting the Tribes: Using Text for Marketing Insight," *Journal of Marketing*, 84(1), 1–25.

Bertacchini, Francesca, Eleonora Bilotta, and Pietro Pantano (2017), "Shopping with a Robotic Companion," *Computers in Human Behavior*, 77, 382–95, https://doi.org/10.1016/j.chb.2017.02.064.

Betancourt, Hector (1990), "An Attribution-Empathy Model of Helping Behavior," *Personality and Social Psychology Bulletin*, 16(3), 573–91.

Birch, S. A.J., V. Li, T. Haddock, S. E. Ghrear, P. Brosseau-Liard, A. Baimel, and M. Whyte (2017), "Perspectives on Perspective Taking: How Children Think About the Minds of Others," *Advances in Child Development and Behavior*, 52, 185–226.

Bretherton, Inge and Marjorie Beeghly (1982), "Talking about Internal States: The Acquisition of an Explicit Theory of Mind," *Developmental Psychology*, 18(6), 906–21.

Bretherton, Inge, Sandra McNew, and Marjorie Beeghly-Smith (1981), "Early Person Knowledge as Expressed in Gestural and Verbal Communication: When Do Infants Acquire a 'Theory of Mind'?," in *Infant Social Cognition*, ed. M. E. Lamb and L. R. Sherrod, Hillsdale, NJ: Erlbaum, 333–73.

Van den Broeck, Evert, Brahim Zarouali, and Karolien Poels (2019), "Chatbot Advertising Effectiveness: When Does the Message Get Through?," *Computers in Human Behavior*, 98(April), 150–57.

Cialdini, Robert B and B. J. Sagarin (1994), "Principles of Interpersonal Influence," in *Persuasion: Psychological Insights and Perspectives*, ed. Timothy C. Brock and Melanie C. Green, Thousand Oaks, CA: Sage Publications, Inc., 143–69.

Columbus, Louis (2020), "What's New In Gartner's Hype Cycle For AI, 2020," *Forbes Magazine*, https://www.forbes.com/sites/louiscolumbus/2020/10/04/whats-new-in-gartners-hype-cycle-for-ai-2020/?sh=75bdd337335c.

Comendador, Benilda E. V., Bien M. B. Francisco, Jefferson S. Medenilla, Sharleen M. T. Nacion, and Timothy B. E. Serac (2015), "Pharmabot: A Pediatric Generic Medicine Consultant

Chatbot," *Journal of Automation and Control Engineering*, 3(2), 137–40, http://www.joace.org/index.php?m=content&c=index&a=show&catid=42&id=218.

Danescu-Niculescu-Mizil, Cristian, Lillian Lee, Bo Pang, and Jon Kleinberg (2012), "Echoes of Power: Language Effects and Power Differences in Social Interaction," *WWW'12 - Proceedings of the 21st Annual Conference on World Wide Web*, 699–708.

Davis, Fred D. (1989), "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology," *MIS Quarterly: Management Information Systems*, 13(3), 319–39.

Deloitte (2019), *Conversational AI: The Next Wave of Customer and Employee Experiences*.

Dietvorst, Berkeley J., Joseph P. Simmons, and Cade Massey (2018), "Overcoming Algorithm Aversion: People Will Use Imperfect Algorithms If They Can (Even Slightly) Modify Them," *Management Science*, 64(3), 1155–70.

Dunn, Judy, Jane Brown, Cheryl Slomkowski, Caroline Tesla, and Lise Youngblade (1991), "Young Children's Understanding of Other People's Feelings and Beliefs : Individual Differences and Their Antecedents," *Child Development*, 62, 1352–66.

Epley, Nicholas, Adam Waytz, and John T. Cacioppo (2007), "On Seeing Human: A Three-Factor Theory of Anthropomorphism," *Psychological Review*, 114(4), 864–86.

Finch, John F. and Robert B Cialdini (1989), "Another Indirect Tactic of (Self-) Image Managment: Boosting," *Personality and Social Psychology Bulletin*, 15(2), 222–32.

Fransen, Marieke L., Edith G. Smit, and Peeter W. J. Verlegh (2015), "Strategies and Motives for Resistance to Persuasion: An Integrative Framework," *Frontiers in Psychology*, 6.

Gächter, Simon, Benedikt Herrmann, and Christian Thöni (2004), "Trust, Voluntary Cooperation, and Socio-Economic Background: Survey and Experimental Evidence," *Journal of Economic Behavior and Organization*, 55(4 SPEC.ISS.), 505–31.

Glikson, Ella and Anita Williams Woolley (2020), "Human Trust in Artificial Intelligence: Review of Empirical Research," *Academy of Management Annals*, 14(2), 627–60.

Go, Eun and S. Shyam Sundar (2019), "Humanizing Chatbots: The Effects of Visual, Identity and Conversational Cues on Humanness Perceptions," *Computers in Human Behavior*, 97(January), 304–16.

Goldstein, Noah J., Robert B. Cialdini, and Vladas Griskevicius (2008), "A Room with a Viewpoint: Using Social Norms to Motivate Environmental Conservation in Hotels," *Journal of Consumer Research*, 35(3), 472–82.

Goldstein, Thalia R and Ellen Winner (2012), "Enhancing Empathy and Theory of Mind," *Journal of Cogntition and Development*, 13(1), 19–37.

Granulo, Armin, Christoph Fuchs, and Stefano Puntoni (2019), "Psychological Reactions to Human versus Robotic Job Replacement," *Nature Human Behaviour*, 3(10).

Gray, Heather M., Kurt Gray, and Daniel M. Wegner (2007), "Dimensions of Mind Perception," *Science*, 315(5812), 619.

Hexmoor, Henry, Shahram Rahimi, and Rachil Chandran (2008), "Delegations Guided by Trust and Autonomy," *Web Intelligence and Agent Systems*, 6(2), 137–55.

Hoff, Kevin Anthony and Masooda Bashir (2015), "Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust," *Human Factors*, 57(3), 407–34.

Hoffman, Donna L. and Thomas P. Novak (2018), "Consumer and Object Experience in the Internet of Things: An Assemblage Theory Approach," *Journal of Consumer Research*, 44(6), 1178–1204.

Hohenberger, Christoph, Chaiwoo Lee, and Joseph F. Coughlin (2019), “Acceptance of Robo-Advisors: Effects of Financial Experience, Affective Reactions, and Self-Enhancement Motives,” *Financial Planning Review*, 2(2).

Holzwarth, Martin, Chris Janiszewski, and Marcus M Neumann (2006), “The Influence of Avatars on Online Consumer Shopping Behavior,” *Journal of Marketing*, 70(4), 19–36.

Humphreys, Ashlee and Rebecca Jen Hui Wang (2018), “Automated Text Analysis for Consumer Research,” *Journal of Consumer Research*, 44(6), 1274–1306.

Jin, Seung A.Annie and Yongjun Sung (2010), “The Roles of Spokes-Avatars’ Personalities in Brand Communication in 3D Virtual Environments,” *Journal of Brand Management*, 17(5), 317–27.

Jurafsky, Daniel and James Martin (2017), “Dialog Systems and Chatbots,” in *Speech and Language Processing*, Pearson, 418–40.

Kim, Youjeong and S. Shyam Sundar (2012), “Anthropomorphism of Computers: Is It Mindful or Mindless?,” *Computers in Human Behavior*, 28(1), 241–50.

Klesse, Anne Kathrin, Jonathan Levav, and Caroline Goukens (2015), “The Effect of Preference Expression Modality on Self-Control,” *Journal of Consumer Research*, 42(4), 535–50.

Landis, Geoffrey A. (2014), “Future Tense: The Chatbot and the Drone,” *Communications of the ACM*, 57(7), 111–12.

Lee, John D. and Katrina A. See (2004), “Trust in Automation: Designing for Appropriate Reliance,” *Human Factors*, 46(1), 50–80.

Lee, Sangwon, Naeun Lee, and Young June Sah (2019), “Perceiving a Mind in a Chatbot: Effect of Mind Perception and Social Cues on Co-Presence, Closeness, and Intention to Use,” *International Journal of Human-Computer Interaction*, 7318.

Levinson, Stephen C. (2016), “Turn-Taking in Human Communication - Origins and Implications for Language Processing,” *Trends in Cognitive Sciences*, 20(1), 6–14.

Lokman, Abbas Saliimi and Jasni Mohamad Zain (2010), “Extension and Prerequisite: An Algorithm to Enable Relations Between Responses in Chatbot Technology,” *Journal of Computer Science*, 6(10), 1212–18.

Longoni, Chiara, Andrea Bonezzi, and Carey K Morewedge (2019), “Resistance to Medical Artificial Intelligence,” *Journal of Consumer Research*, 46, 629–50.

Luo, Xueming, Siliang Tong, Zheng Fang, and Zhe Qu (2019), “Frontiers: Machines vs. Humans: The Impact of Artificial Intelligence Chatbot Disclosure on Customer Purchases,” *Marketing Science*, 38(6), 937–47.

Maner, Jon K., Carol L. Luce, Steven L. Neuberg, Robert B. Cialdini, Stephanie Brown, and Brad J. Sagarin (2002), “The Effects of Perspective Taking on Motivations for Helping: Still No Evidence for Altruism,” *Personality and Social Psychology Bulletin*, 28(11), 1601–10.

McAllister, Daniel J. (1995), “Affect and Cognition-Based Trust as Foundations for Interpersonal Cooperation in Organizations,” *Academy of Management Journal*, 38(1), 24.

Melumad, Shiri, Rhonda Hadi, Christian Hildebrand, and Adrian F. Ward (2020), “Technology-Augmented Choice: How Digital Innovations Are Transforming Consumer Decision Processes,” *Customer Needs and Solutions*.

Mende, Martin, Maura L. Scott, Jenny van Doorn, Dhruv Grewal, and Ilana Shanks (2019), "Service Robots Rising: How Humanoid Robots Influence Service Experiences and Elicit Compensatory Consumer Responses," *Journal of Marketing Research*, 56(4), 535–56.

Miner, Adam, Amanda Chow, Sarah Adler, Ilia Zaitsev, Paul Tero, Alison Darcy, and Andreas Paepcke (2016), "Conversational Agents and Mental Health: Theory-Informed Assessment of Language and Affect," *HAI 2016 - Proceedings of the 4th International Conference on Human Agent Interaction*, 123–30.

Moon, Youngme (2000), "Intimate Exchanges: Using Computers to Elicit Self-Disclosure From Consumers," *Journal of Consumer Research*, 26(4), 323–39.

Nass, Clifford and Youngme Moon (2000), "Machines and Mindlessness: Social Responses to Computers," *Journal of Social Issues*, 56(1), 81–103.

Nass, Clifford, Youngme Moon, B. J. Fogg, Byron Reeves, and D. Christopher Dryer (1995), "Can Computer Personalities Be Human Personalities?," *International Journal of Human - Computer Studies*, 223–39.

Nass, Clifford, Jonathan Steuer, and Ellen R. Tauber (1994), "Computers Are Social Actors," *Conference companion on Human factors in computing systems - CHI '94*, 204.

Novak, Thomas P. and Donna L. Hoffman (2019), "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), 216–37.

Oh, Kyo Joong, Dongkun Lee, Byungsoo Ko, and Ho Jin Choi (2017), "A Chatbot for Psychiatric Counseling in Mental Healthcare Service Based on Emotional Dialogue Analysis and Sentence Generation," *Proceedings - 18th IEEE International Conference on Mobile Data Management, MDM 2017*, 371–76.

Paul, Elizabeth S., Anna Moore, Pippa McAinsh, Emma Symonds, Sandra McCune, and John W.S. Bradshaw (2014), "Sociality Motivation and Anthropomorphic Thinking about Pets," *Anthrozoos*, 27(4), 499–512.

Premack, David and Guy Wooclryff (1978), "Does the Chimpanzee Have a Theory of Mind," *The Behavioral and Brain Sciences*, 4, 515–26.

Rousseau, Denise M., Sim B. Sitkin, Ronald S. Burt, and Colin Camerer (1998), "Not so Different After All: A Cross-Discipline View of Trust," *Academy of Management Review*, 23(3), 393–404.

Schmitt, Bernd (2019), "From Atoms to Bits and Back: A Research Curation on Digital Technology and Agenda for Future Research," *Journal of Consumer Research*, 46(4), 825–32.

Schroeder, Juliana, Michael Kardas, and Nicholas Epley (2017), "The Humanizing Voice: Speech Reveals, and Text Conceals, a More Thoughtful Mind in the Midst of Disagreement," *Psychological Science*, 28(12), 1745–62.

Serban, Iulian V., Chinnadhurai Sankar, Mathieu Germain, Saizheng Zhang, Zhouhan Lin, Sandeep Subramanian, Taesup Kim, Michael Pieper, Sarath Chandar, Nan Rosemary Ke, Sai Rajeshwar, Alexandre de Brebisson, Jose M. R. Sotelo, Dendi Suhubdy, Vincent Michalski, Alexandre Nguyen, Joelle Pineau, and Yoshua Bengio (2017), "A Deep Reinforcement Learning Chatbot," *ArXiv preprint*, 1–34.

Sheehan, Ben, Hyun Seung Jin, and Udo Gottlieb (2020), "Customer Service Chatbots: Anthropomorphism and Adoption," *Journal of Business Research*, 115(April), 14–24.

Sundar, Shyam S. and Clifford Nass (2000), "Source Orientation in Human-Computer Interaction: Programmer, Networker or Independent Social Actor?," *Communication Quarterly*, 27(6), 683–703.

Tapus, Adriana and Maja J. Mataric (2008), “Socially Assistive Robots: The Link between Personality, Empathy, Physiological Signals, and Task Performance,” *AAAI Spring Symposium - Technical Report*, SS-08-04, 133–40.

Tedesco, Darren (2015), “I, Robo-Adviser? Creating the Blended Adviser Experience.,” *Journal of Financial Planning*, 28(1), 17–19.

Turing, Alan M. (1950), “Computing Machinery and Intelligence,” *Mind*, 59(236), 433–60.

Wang, Liz C, Julie Baker, Judy A Wagner, and Kirk Wakefield (2007), “Can a Retail Web Site Be Social?,” *Journal of Marketing*, 71(3), 143–57.

Waytz, Adam, John Cacioppo, and Nicholas Epley (2010), “Who Sees Human? The Stability and Importance of Individual Differences in Anthropomorphism,” *Perspectives on Psychological Science*, 5(3), 219–32.

Waytz, Adam, Kurt Gray, Nicholas Epley, and Daniel M. Wegner (2010), “Causes and Consequences of Mind Perception,” *Trends in Cognitive Sciences*, 14(8), 383–88.

Yang, Linyun W, Pankaj Aggarwal, and Ann L McGill (2020), “The 3 C’s of Anthropomorphism: Connection, Comprehension, and Competition,” *Consumer Psychology Review*, 3(1), 3–19.

Zarouali, Brahim, Evert Van Den Broeck, Michel Walrave, and Karolien Poels (2018), “Predicting Consumer Responses to a Chatbot on Facebook,” *Cyberpsychology, Behavior, and Social Networking*, 21(8), 491–97.

# 2. ARTICLE I

Bergner, A., Hildebrand, C., and Häubl, G. (under review: third round). Machine Talk: How AI-Based Conversational Interfaces Enhance Brand Intimacy and Impact Behavioral Brand Outcomes. *Journal of Consumer Research.*

**Abstract**

AI-based conversational interfaces are often thought of merely as tools for service automation to provide 24/7 customer support. This research shows that such interfaces can also have a profound impact on consumer-brand relationships. We introduce a conceptual model of technology-mediated communication that identifies the structural (turn-taking, turn-initiation) and semantic (grounding) properties of a dialogue as the key drivers of consumers' perception of an AI-based conversational interface, their perception of the brand that the interface represents, and their behavior in connection with the brand. Converging evidence from five studies shows that these conversational properties profoundly impact the perceived humanness of conversational interfaces, which promotes more intimate consumer-brand relationships, ultimately resulting in more favorable behavioral brand outcomes (such as greater recommendation acceptance, willingness to pay a price premium, brand advocacy, and brand loyalty). Moreover, we show that these effects are enhanced in contexts requiring greater mutual understanding between the consumer and the brand. This research highlights how our understanding of fundamental principles of human-to-human communication can be harnessed to design more relatable AI. The findings presented here have important practical implications for the design of AI-based conversational interfaces that brands use to interact with consumers.

*Keywords*: Chatbots, Conversational Interfaces, Natural Language Processing, Brand Intimacy, Consumer-Brand Relationships, Technology-Mediated Branding.

# Machine Talk: How AI-Based Conversational Interfaces Enhance Brand Intimacy and Impact Behavioral Brand Outcomes

Advances in natural language processing and the increased adoption of messaging platforms are transforming how consumers interact with brands (BI Intelligence 2016; Dale 2016; Hirschberg and Manning 2015). The use of AI-based conversational interfaces as an interaction paradigm between consumers and brands has been characterized as the "next operating system in commerce" (Suri, Elia, and van Hillegersberg 2017; The Economist 2016). Conversational interfaces are becoming an increasingly integral part of brands' digital strategies as consumers shy away from traditional websites in favor of more immediate messenger systems they already use to interact with their friends, and now also with firms (Heikes 2017). Conversational interfaces (also referred to as "chatbots" or "conversational AI") allow brands to engage consumers in an immediate, two-way conversation, and to provide information about products and services to consumers (Clark 2017). Retailers such as Sephora have adopted chatbots to provide virtual product consulting, shopping assistance, and scheduling support for in-store appointments (Johnston Taylor 2016). While recent industry reports indicate that conversational interfaces are used by at least 25% of all customer-service and support operations (Gartner 2018), the specific implementation of these interfaces varies substantially, ranging from the WHO's simple FAQ bot for accessing health information to Starbuck's Barista bot that engages in active, two-way dialogue with the consumer to take orders or provide recommendations (see Web Appendix A1 for additional examples).

Academic research on AI-based conversational interfaces has been predominantly concerned with design and technology-related factors. Specifically, the majority of prior work on conversational interfaces examined how firms might optimize the design features of conversational interfaces (Landis 2014; Lokman and Zain 2010), factors related to user perceptions and acceptance (Comendador et al. 2015; Go and Sundar 2019; Hildebrand and Bergner 2020; Lee et al. 2019; Zarouali et al. 2018), user attitudes (Araujo 2018; Bellur and Sundar 2017), or how conversational interfaces might learn in an unsupervised fashion through interactions with a user (Serban et al. 2017). Research on how conversational interfaces impact consumers' brand perceptions, how these conversations should be "engineered," and the downstream impact on consumer decision making has been scarce and primarily focused on the influence of presenting avatars or other visual characteristics of embodiment on consumers (Bertacchini, Bilotta, and Pantano 2017; Holzwarth, Janiszewski, and Neumann 2006; Jin and Sung 2010; Wang et al. 2007).

In the current research, we develop a rigorous conceptualization of brands' AI-based conversational interfaces integrating prior research on consumer-brand relationships as an active social process (Brick and Fournier 2017; Fournier 1998; MacInnis and Folkes 2017; Thomson 2006) with fundamental principles of human-to-human dialogue (Grice 1975; Jurafsky and Martin 2017; Searle 1969; Wiemann and Knapp 1975). We introduce a theoretical framework that characterizes how verbal embodiment of key properties of human dialogue in a conversational interface influences consumers' perception of the interface and of the brand it represents, as well as important downstream behavioral brand outcomes. According to our theorizing, the structural and semantic properties that govern the process of a conversation (turn-taking, turn-initiation, and grounding) have a positive impact on the perceived humanness of a brand's conversational interface, in turn promoting a more intimate relationship with the brand, and ultimately causing consumers to engage in more favorable brand-related behaviors (such as greater brand loyalty, brand advocacy, and a greater willingness to pay a price premium).

In what follows, we first develop a theoretical framework for understanding how the nature of a brand's AI-based conversational interface impacts consumers' perception of, and behavior related to, the brand. We then present evidence from five studies that were designed to test this theorizing. We conclude with a discussion of how this work contributes to the emerging body of research on technology-mediated consumer-brand interactions, the theory-driven design of effective conversational interfaces in industry practice, and the broader implications of the findings for the future of branding and more relatable AI.

## Theory and Hypotheses

### *Verbal Embodiment in Conversational Interfaces*

Language is commonly considered "the mark of humanity and sentience" and dialogue is thought to represent "the most fundamental and specially privileged arena of language" (Jurafsky and Martin 2017). With the proliferation of AI-based conversational interfaces, machines are entering a territory of natural language that was previously exclusive to humans. Research suggests that human users often fail to realize that a chatbot they are "conversing" with is not a human (Metz 2020). Based on the concept of an "embodied system" as a computing interface that mimics a human (Cassell et al. 2000), we posit that a conversational interface becomes an embodied system by utilizing fundamental principles of human dialogue. We further propose that such an embodied system operates not through visual embodiment (i.e., human face and body) but through the embodiment of key properties of human dialogue.

Conversation is a collective activity in which two interacting parties, instead of merely issuing utterances independently, must coordinate on both process and content (Clark and Brennan 1991; Grice 1989). In this research, we focus on the essential properties that shape the process of a conversational interaction, disregarding any other design aspects of the interface (e.g., a humanized avatar) or content of the conversation (e.g., emojis, typeface, formality of the language) that might further influence embodiment (Li, Chan, and Kim 2019). Building on Cassell et al.'s (2000) definition of embodied systems, we conceptualize "verbal embodiment" in an automated conversational interface as the AI-based system's act of mimicking specific properties of human-to-human dialogue.

Human conversation is fundamentally governed by two general principles of process coordination – (1) how the flow of the conversation is structured between two entities (i.e., the amount of turn-taking and how turns are initiated) and (2) the degree to which the conversation entails semantic grounding in that mutual understanding is expressed and affirmed as part of the dialogue (see Figure 1 below for a schematic overview of these structural and semantic properties). These communication principles are what makes conversations fundamentally social, co-constructed activities (Clark and Brennan 1991; Clark and Schaefer 1989; Searle 1969).

*Turn-Taking.* A fundamental property of human-to-human conversations is the notion of turn-taking – the knowledge of when to speak, and what it means when someone interrupts or remains silent (op den Akker and Bruijnes 2012), making dialogue a cooperative, social act (Grice 1975). Turn-taking among humans is universal across languages. It requires significant cognitive resources, and it stems from a combination of instinctive and learned behaviors (Levinson 2016). Infants as young as six months try to mimic structural properties in a "conversation" even though they do not understand the spoken message content, while around the age of nine months, turn-taking becomes more synchronized (Levinson 2016). Hence, rather than simply a question of recognizing and acting on the behavioral cues of one's conversational partner, turn-taking is a "truly responsive or co-constructed activity" (Cassell et al. 2001), reflecting a uniquely human cognitive ability.

Turn-taking has important implications as it helps conversational partners shape their relationship to one another, as well as define their respective roles during an interaction (Wiemann and Knapp 1975). For example, Sprecher et al. (2013) demonstrated that following a turn-taking protocol produces more favorable interpersonal outcomes compared to conversational partners not engaging in any turn-taking. Specifically, they showed that participants who disclosed reciprocally, as opposed to through a non-reciprocal monologue, reported greater liking, closeness, perceived similarity with the conversational partner, and enjoyment of the interaction.

Prior work on embodied systems also revealed that human users tended to prefer interfaces that are capable of turn-taking and experienced greater comfort during an interaction when the interface engaged in more turn-taking (Cassell et al. 2001). Indeed, research on embodiment in information systems demonstrated that representing the structural process of a dialogue through embodiment (e.g., using visual feedback such as a nod or eye contact to give and take turns during a conversation) is more important than the actual display of emotions (e.g., smiling at the user), and had a greater impact on making the computing interface appear more human-like and helpful (Cassell and Thórisson 1999). Against this background, we hypothesize that the capacity of an AI-based conversational interface to mimic human dialogue through a reciprocal process of turn-taking promotes consumers' attributions of humanness toward that interface.

**$H_1$:** Turn-taking between a consumer and a brand's AI-based conversational interface enhances the perceived humanness of the interface.

*Turn-Initiation.* The second structural aspect of conversational process coordination is turn-initiation, which is sometimes also referred to as "conversational initiative" (Cassell et al. 2000; Jurafsky and Martin 2017; Walker and Whittaker 1995). Turn-initiation can be thought of as the extent of control that each party has over a conversational interaction that entails turn-taking. More specifically, if a conversational interface can only directly respond to a question posed by the user, then the interaction is fully controlled by the user. In human-to-human dialogue, it is more common for both parties to be able to initiate turns, such as by answering a question and then immediately initiating a turn by asking a follow-up question (Jurafsky and Martin 2017).

The ability to initiate turns can also be thought of as reflecting the extent to which a system can control its own actions, which affects the perceived "socialness" of an interface (Nagao and Takeuchi 1994) and may foster a sense of cooperation during a conversation. Turn-initiation is an important design feature of AI-based conversational interfaces and can range from very low levels (e.g., similar to a FAQ bot, during which the user fully controls the interaction and the interface only reacts to the input of the user) to high levels (e.g., similar to a service bot, where the conversational interface guides the interaction through a set of questions and autonomously responds to user input).

Prior work in social psychology has shown that people who either have, or are at least perceived as having, greater control over their own actions are not only liked more, but also tend to be preferred as conversation or interaction partners (Deci and Flaste 1995). Moreover, recent qualitative work on the reasons why consumers often prefer "smart" products over traditional, less "intelligent" products (Wünderlich, Wangenheim, and Bitner 2013) revealed that, despite a general preference for convenience (Brown 1989; Yale and Venkatesh 1986), a key driver of consumer preference for smart products is that they adapt themselves to the human user (instead of requiring the user to adapt to the system), creating the perception of a more cooperative relationship between the consumer and the product.

As human-to-human dialogue generally allows for both parties to initiate turns during a conversation, we hypothesize that the capacity of a brand's AI-based conversational interface to initiate a turn during an interaction process that entails turn-taking promotes consumers' attributions of greater humanness to that interface.

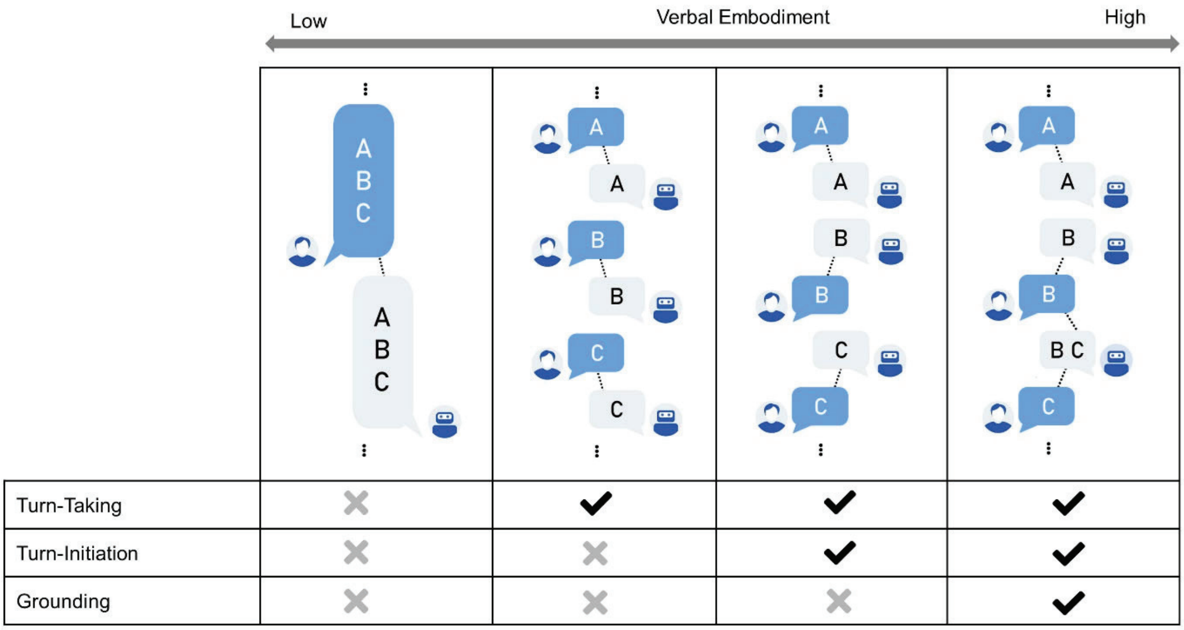
**$H_2$:** Turn-initiation by a brand's AI-based conversational interface enhances the perceived humanness of the interface.

*Grounding*. While turn-taking and turn-initiation facilitate the structural organization of a conversation, grounding further enhances this process by adding signals of mutual understanding to individual turns (Clark and Brennan 1991). Grounding is the act of establishing a “common ground” by acknowledging that the listener has understood the speaker (Jurafsky and Martin 2017) through both verbal (e.g., “okay”) and non-verbal (e.g., nodding or eye-movement) acknowledgements (see Richardson and Dale 2004). Verbal acknowledgements can be made more explicit by repeating or paraphrasing what the speaker has said (McInnes and Attwater 2004). Thus, verbal grounding does not add new content to a conversation but it is a semantic mechanism to acknowledge that the listener has received and understood the speaker’s contribution to the conversation. By grounding each other’s contributions, conversational responses become interconnected and contingent on what has been said previously (Sundar, Qian, and Bellur 2010), a key characteristic of human discourse (Ghose and Barua 2013).

Moreover, the act of grounding a message by acknowledging or paraphrasing it confirms to the speaker that the other party is actively listening to what s/he is saying. Active listening has been shown to be an important communication skill during initial interactions, suggesting that the other party is interested in, and paying attention to, what is being said, and thus fostering the impression of mutual understanding (Weger et al. 2014). These attributions can have important consequences for the relationship perception of the interacting parties. For instance, Ramsey and Sohi (1997) found that customers’ perception of a salesperson’s listening behavior played an important role in trust formation and as an antecedent of consumers’ inclination to interact with the same salesperson again in the future. Thus, we hypothesize that, in the context of a brand’s conversational interface, grounding leads to greater attributions of humanness to that interface.

**$H_3$:** Grounding between a brand’s AI-based conversational interface and a consumer enhances the perceived humanness of the interface.

Figure 1 provides a schematic overview of the structural (turn-taking and turn-initiation) and semantic (grounding) properties that define the conversational process during a consumer-brand interaction mediated by an AI-based conversational interface.




*Note*: Letters represent topics. Dots at the top and bottom indicate that this is merely a portion of the interaction – i.e., there could be communication before and after what is illustrated here.

**Figure 2-1**: Key conceptual properties of AI-based conversational interfaces

*Leveraging Turn-Taking, Turn-Initiation, and Grounding to Enhance Brand Intimacy*

In human-to-human relationships, the formation of intimacy is a critical social process that serves as a foundation for a sense of mutual understanding between two entities (Reis & Shaver, 1988; Clark & Reis, 1988). The fact that greater intimacy in relationships is often characterized by more extensive self-disclosure (Sprecher and Hendrick 2004) and reciprocity (McAdams, Jackson, and Kirshnit 1984) supports the idea that intimacy is closely associated with the formation and deepening of interpersonal connections (Reis and Shaver 1988). Thus, intimacy includes, among other things, a concern for the welfare of, high regard for, and intimate communication with, the other entity (Sternberg and Grajek 1984). Moreover, it is a critical antecedent of attachment to another entity (Baumeister and Leary 1995; Deci and Ryan 2000; La Guardia et al. 2000) and of commitment to a relationship (Gilbert 1976).

In her seminal work on brands as relationship partners, Fournier (1998) proposed that consumer-brand relationships are based on similar social processes as interpersonal relationships. Thus, the perceived intimacy with a brand is reflected in the apparent closeness and extent of mutual understanding between the consumer and the brand (Thorbjørnsen et al.

2002), and it has been shown to be an important antecedent of other brand-related outcomes such as brand love (Carroll and Ahuvia 2006; Rauschnabel and Ahuvia 2014) and brand attachment (Thomson, MacInnis, and Park 2005b). The construct of brand intimacy captures the degree to which consumers experience a (developing) bond with the brand and feel that the brand understands and cares about them (Thorbjørnsen et al. 2002). Thus, whereas the focus of other brand-related constructs such as self-brand connection or brand attachment is on the extent to which the brand is part of a consumer's sense of "self" or expresses who they are (Escalas 2004; Park et al. 2010), brand intimacy specifically refers to the extent of understanding between the consumer and the brand (Fournier 1998), as perceived by the consumer, which results in the consumer feeling cared about and validated by the brand (Natarelli and Plapler 2017; Thorbjørnsen et al. 2002).

Based on this prior work in domains ranging from interpersonal relationships to branding, as well as recent research showing that embodied gestures can lead to more intimacy with a brand (Hadi and Valenzuela 2014), we propose that verbal embodiment is a key mechanism through which a brand's AI-based conversational interface can establish a closer connection with the consumer. Specifically, turn-taking in a conversation signals cooperation and serves as the basic requirement for a dynamic interaction between two (or more) parties. Turn-taking without turn-initiation implies a one-sided dialogue, whereas turn-initiation further signals active participation by the other party. Verbal grounding entails references to the content of previous turns and thus enhances mutual understanding by establishing connections among individual turns. These verbal embodiments in a brand's conversational interface should facilitate the development of a deeper bond between the consumer and the brand. Thus, we hypothesize that interactions with an AI-based interface that is perceived as more human-like, as a result of its ability to take and initiate turns and to increase mutual understanding through grounding, promote more intimate consumer-brand relationships.

**$H_4$:** The perceived humanness of a brand's AI-based conversational interface has a positive effect on consumers' experienced brand intimacy.

*The Impact of Enhanced Brand Intimacy on Behavioral Brand Outcomes*

Greater intimacy in human-to-human relationships is a key driver of interpersonal influence (Cialdini and Sagarin 1994; Finch and Cialdini 1989; Goldstein et al. 2008; Maner et al. 2002). That is, we are more likely to comply with requests of individuals we like and feel close to. Work by Holzwarth, Janiszewski, and Neumann (2006) suggests that a similar phenomenon might occur when humans interact with humanized avatars. Their findings demonstrate that the mere presence of such avatars can help reduce the impersonal nature of web-based shopping interfaces, and that the personality and attractiveness of the avatar can further influence the persuasiveness of its messages. Related work by Gong (2008) suggests that maximizing the human appearance of an interface can increase users' willingness to comply with requests. Although the psychological mechanisms underlying these phenomena have not been fully unveiled, the observed effects do suggest the possibility that perceptions of the humanness of conversational interfaces might, via feelings of intimacy with the brands they represent, affect behavioral brand outcomes.

Thus, we hypothesize that the level of brand intimacy that consumers perceive as a consequence of interacting with a brand via its AI-based conversational interface has a positive impact on key behavioral outcomes that are desirable from the brand's perspective. The brand-related behaviors that we examine in this research range from ones that are specific to the transaction that is the focus of the interaction – consumers' acceptance of a recommendation made by the brand and their willingness to pay a price premium for the brand's offerings – to behaviors that go beyond the current transaction – consumers' enduring commitment to the brand in the form of brand advocacy and their intention to engage in future transactions (i.e., brand loyalty).

**$H_5$:** Greater brand intimacy has a positive effect on consumers' (1) inclination to accept brand recommendations, (2) willingness to pay a price premium for the brand's products, (3) motivation to advocate the brand to others, and (4) brand loyalty.

*The Moderating Role of Context-Specific Need for Mutual Understanding*

Against the background of our conceptualization of brand intimacy as a co-created social construct, we propose that not all contexts are equally conducive to consumers developing such a sense of intimacy based on their interactions with a brand. In particular, the settings where consumer-brand interaction occurs differ in the extent to which a deep, nuanced understanding of the other party is necessary for such an interaction to result in a mutually satisfying outcome. Similar to the signaling of credibility to reduce consumer uncertainty (Erdem and Swait 1998), a brand may wish to signal that it is motivated to understand individual consumers, particularly under some market conditions. Product choice settings that are inherently more complex due to a greater number of options tend to require a higher level of mutual understanding between a brand (with an expansive array of product offerings) and a consumer (with an idiosyncratic preference; see Dellaert and Stremersch 2005; Hildebrand, Häubl, and Herrmann 2014). For instance, configuring one's preferred type of latte from an extensive, complex menu of options at a high-end espresso shop can require a significant amount of understanding between the vendor and the consumer in order to establish a match between the consumer's idiosyncratic preference and what the vendor might be able to produce, whereas picking up a standard cup of black coffee at a gas station typically does not.

We theorize that this characteristic of the context in which consumer-brand interaction takes place is a key moderator of the mechanism through which the nature of the interface that the brand provides for this interaction shapes the consumer-brand relationship. Specifically, we hypothesize that the conversational properties of a brand's AI-based customer interface tend to have a more pronounced impact on the consumer's experience of brand intimacy in settings where the need for mutual understanding between the brand and the consumer is greater.

**$H_6$:** The positive effects of the conversational properties of a brand's AI-based customer interface on brand intimacy are amplified in settings that require higher levels of mutual understanding.

The conceptual model shown in Figure 2 provides a concise, high-level summary of our theorizing by depicting the hypothesized relationships among the key constructs.

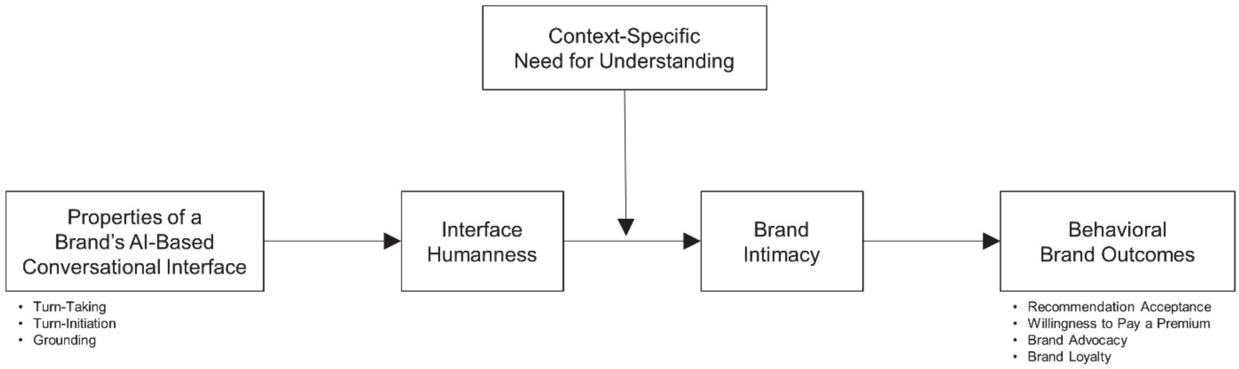



**Figure 2-2:** Conceptual model

## Overview of Studies & Experimental Setting

We present evidence from a pilot study and four main studies that were designed to test our theorizing. The pilot study shows that conversational interfaces are perceived as significantly more human-like than functionally equivalent non-conversational interfaces (i.e., traditional websites), and that they can increase consumers' willingness to pay for a product. Studies 1 to 4 focus on the effects of specific interface properties within AI-based conversational interfaces and on how they shape consumer-brand relationships. Studies 1 and 2 show that the properties of turn-taking and turn-initiation are critical to enhancing the perceived humanness of a brand's interface, and that greater interface humanness in turn increases perceived brand intimacy (Study 2). Study 3 demonstrates that, in addition to the structural aspects of turn-taking and turn-initiation (Studies 1 and 2), the semantic aspects of the conversation in the form of acts of grounding also enhance the perceived humanness of a brand's interface. Study 4 shows that the positive effects of the conversational properties of a brand's AI-based interface on brand intimacy are amplified in settings that require higher levels of mutual understanding. Finally, across studies, the results also demonstrate that the increase in brand intimacy caused by the conversational properties of AI-based interfaces has significant brand-specific downstream consequences (increased recommendation acceptance, willingness to pay a price premium, brand advocacy, and brand loyalty).

In all experiments, we used custom-developed conversational interfaces to ensure maximal experimental control and a realistic study setting involving actual behavior to isolate the proposed effects of both structural (turn-taking, turn-initiation) and semantic (grounding) properties governing the interaction process. As illustrated in the Web Appendix (see Figure

A2.1 in the Web Appendix), we performed regular expression matching to validate the natural language input provided by users across all our studies and ensure both high data quality and realism across experiments (i.e., avoiding that participants can input any data that would be invalid, given a question posed by the conversational interface). All data streams were stored and merged on a separate web server for final analyses. The experimental setup is illustrated and described further in the Web Appendix.

## Pilot Study

We conducted a pilot study to assess whether interacting with a conversational interface as opposed to a non-conversational interface leads to different attributions of human-likeness and the downstream consequences for consumers. This pilot study is based on a field experiment in the fashion industry aimed at testing whether conversational as opposed to non-conversational interfaces impact consumers' valuation of a target product conditional on the type of interface used to customize this target product.

*Design and Procedure.* We designed a field experiment around a fictional dress shirt startup called "SwissThread." Participants were led to believe that they were participating in the company's pre-launch market test. A booth was set-up in the public area of a major Swiss University, including branded banners with SwissThread's logo and a call for participation in a market test. Participants were approached and asked if they would be willing to participate in exchange to win a dress shirt of their preferred design. Those participants who agreed were randomly assigned to either a conversational interface or a non-conversational interface condition (see Figure 3) and given a tablet device to customize a shirt. Each participant was led through the shirt customization process, answering questions on their body type, preference for shirt fit, hair and eye color, along with the customization of their preferred shirt type, color, and style before receiving a recommended shirt option (see Web Appendix A3.1 for details). All questions were identical and presented in the same order in both interface conditions. As shown in Figure 3, in the conversational interface condition, the participants were led through the questions by a conversational interface as if they were having a messenger-based conversation with the brand (right image), while in the non-conversational condition, participants were led through the questions in a standard questionnaire format on a traditional website (left image). In both conditions, participants indicated their answers to each question by selecting from the options provided by the interface. A total of 252 males at a major Swiss university participated in this field experiment. Six participants who clearly did not understand the study instructions were removed, leaving 246 participants for the analyses.

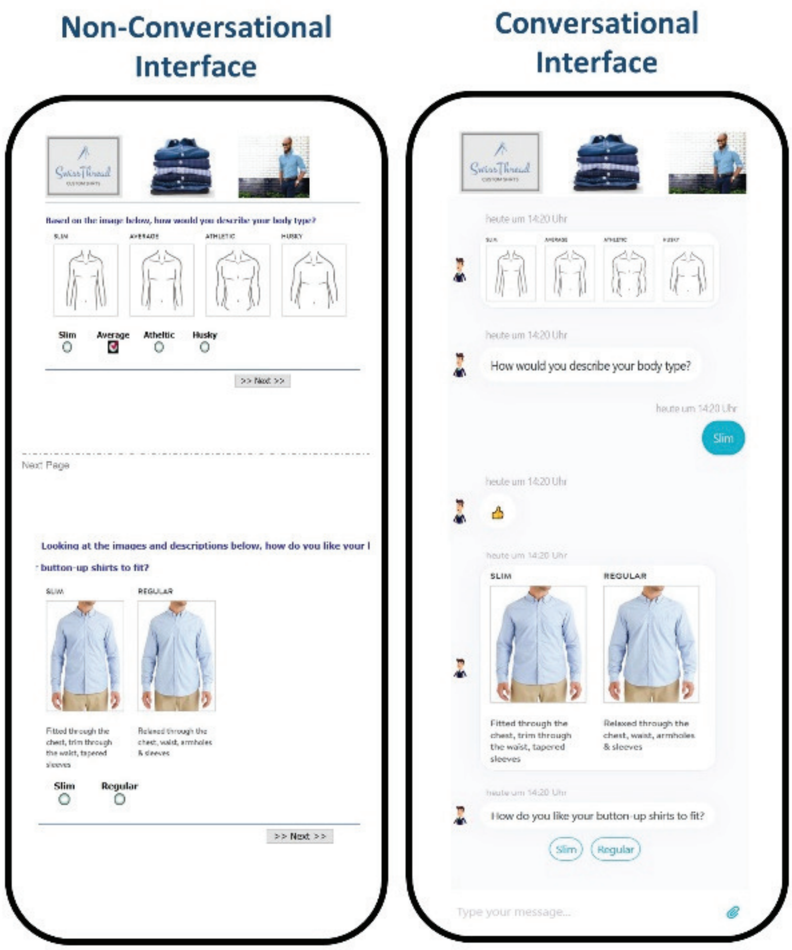



**Figure 2-3:** Interface conditions in the field experiment

*Measures.* We used a Vickrey auction (Noussair, Robin, and Ruffieux 2004) immediately after the shirt customization task to assess differences in product valuation (sealed bid auction; highest bidder wins and price paid is the second-highest bid). Specifically, participants were asked to provide a bid amount on their customized shirt for a chance to win it. Bids were submitted in Swiss Francs. After that, participants rated the perceived humanness of the interface (single-item based on Rijsdijk, Hultink, and Diamantopoulos (2007), "How human did you perceive this interface to be?," 5-point scale).

*Results.* In line with our theorizing, the conversational interface was perceived as significantly more human-like compared to the non-conversational interface ($M_{Conv}$ = 3.47,

$M_{NonConv}$ = 2.86; $F(1, 244) = 20.41$, $p = .001$, $\eta^2 = .08$). Furthermore, participants in the conversational interface condition also bid a marginally higher price for their preferred shirt compared to the non-conversational interface condition ($M_{Conv}$ = CHF 38.37, $M_{NonConv}$ = CHF 34.02; $F(1, 244) = 1.739$, $p = .083$, $\eta^2 = .01$). To further assess our theorizing, we estimated a simple mediation model (5000 bootstrap samples) with the interface manipulation as the independent, perceived humanness as the mediator, and consumers' bid amount as the dependent variable. Mirroring the main effect findings, the conversational interface led to a significant increase in perceived humanness ($\beta_{Condition} = .61$, $t = 4.52$, $p < .001$) and greater perceived humanness led to a significant increase bid amount ($\beta_{Human} = 3.37$, $t = 2.863$, $p < .01$), switching off the effect of the experimental manipulation ($p = .37$) and leading to a significant indirect effect with a confidence interval excluding zero ($\beta_{Indirect} = 2.04$, 95% CI: [.45; 3.98]), indicating full mediation.

*Discussion.* The findings of this pilot field study demonstrate that conversational as opposed to non-conversational interfaces cause consumers to attribute significantly greater levels of human-likeness and subsequently greater willingness to pay for the same branded product. However, the key question is to which extent these findings are driven by the mechanism of the underlying conversational properties of the interface (such as the extent of turn-taking, turn-initiation, or grounding). In what follows, we isolate these conversational properties to test our theorizing on how conversational interfaces impact consumer-brand relationships.

**Study 1**

Study 1 examined the impact of the two key structural properties of a brand's AI-based conversational interface – turn-taking between the consumer and the interface and turn-initiation by the interface – on the perceived humanness of that interface. Based on our theorizing, both the ability to take turns during the interaction with the interface ($H_1$) and the interface's ability to autonomously initiate a turn ($H_2$) should increase the perceived humanness of the interface.

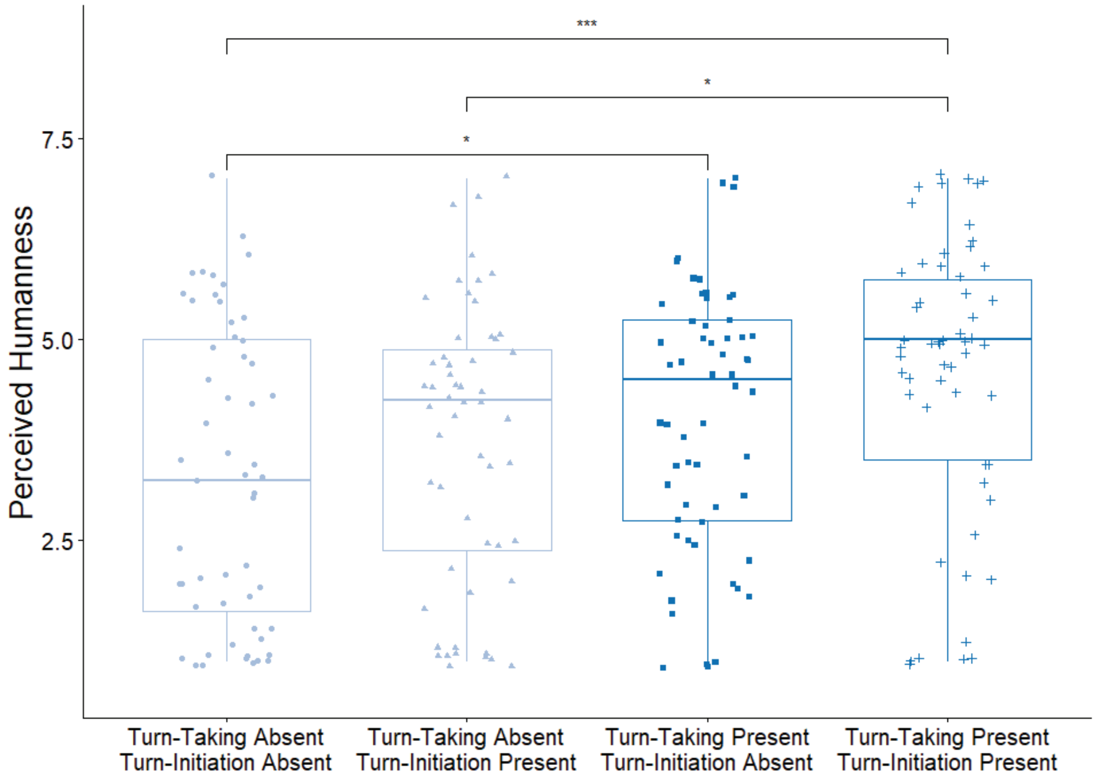
*Design and Procedure.* A total of 226 consumers from MTurk participated in this study ($M_{Age} = 30.89$, $SD_{Age} = 15.00$; 59% females). They were randomly assigned to a 2 (presence vs. absence of turn-taking) × 2 (presence vs. absence of turn-initiation) between-subjects experimental design. Participants were invited to take part in a study on consumer design preferences and informed that they would be interacting with an automated conversational interface. We created a fictitious design brand called “AI.Design” and participants were led to believe that the purpose of the study was to better understand consumers’ design preferences for everyday objects. Next, participants completed a simple design task in which they chose between one of three feature levels for each of eight design features (e.g., geometric shape, fill pattern, line pattern, ornaments etc.; see Web Appendix A3.2 for details).

In the conditions with turn-taking, the interface first provided participants all three feature levels and then asked participants to indicate their preferred feature level by typing their answer into the chat console at the bottom of the interface. After participants provided their answer, the system went on with the next feature and continued sequentially until all feature-levels were specified. In the conditions without turn-taking, the interface provided identical features and feature-levels in exactly the same order but all options were presented simultaneously. After the interface presented all features and feature levels, participants then specified their preferred feature level by typing their answer into the chat console.

The turn-initiation manipulation was conducted as follows. In the turn-initiation present condition, the interface directly performed the next question in the sequence of feature level evaluations, thus playing an active role in the interaction. By contrast, in the turn-initiation absent condition, the interface had a passive role and participants controlled the flow of the interaction through clicking on a “Next” button to see the next feature and corresponding feature levels. Note that this manipulation did not affect participants’ input to the design question, and they were not able to select a different option once they had made a selection and before clicking on the “Next” button.

*Measures.* Immediately after participants completed the design task, we assessed the perceived humanness of the interface (Rijsdijk et al. 2007) using four items ( “This configuration system …has human properties, …is like a person, …behaves like a human being, …has its own character”, 7-point Likert scale, from 1:“I do not agree at all” to 7:“I fully agree”; $\alpha_{human}$ = .96; see Web Appendix A5.1). Next, we used a qualitative, open response technique to assess consumers’ spontaneous brand-related thoughts. Participants first wrote down their spontaneous thoughts or feelings related to the “AI.Design” brand and then evaluated the valence of each of these thoughts as negative (coded as -1), neutral (coded as 0), or positive (coded as 1). In the following analyses, all text entries of the thought-listing task were concatenated into a single text vector. We used the AFINN lexicon (Nielsen 2011) for all subsequent sentiment analyses to evaluate consumers’ brand sentiment conditional on our experimental manipulations.

*Results.* In support of our theorizing, we find that both turn-taking ($H_1$) and the interface’s ability to initiate turns in a conversation ($H_2$) significantly influenced how participants perceived the interface. As illustrated in Figure 4, turn-taking significantly increased perceived humanness ($M_{TurnTaking\_Present} = 4.28$, $M_{TurnTaking\_Absent} = 3.52$; $F(2, 222) = 10.561$, $p < .01$, $\eta^2 = .05$). Furthermore, turn-initiation by the system further increased the level of perceived humanness ($M_{TurnInitiation\_Present} = 4.15$, $M_{TurnInitiation\_Absent} = 3.68$; $F(2, 222) = 4.137$, $p < .05$, $\eta^2 = .02$). The interaction between the two factors was non-significant ($F(1, 222) = 0.06$, $p > .80$), indicating that both properties act as independent forces to increase consumers’ perception of humanness.




*Note:* *** = $p < .001$, ** = $p < .01$, * = $p < .05$;

**Figure 2-4:** Turn-taking and turn-initiation as drivers of humanness perceptions

Next, we further explored participants' spontaneous thoughts about the "AI.Design" brand. In line with our theorizing and mirroring the effects on perceived humanness, both turn-taking ($M_{TurnTaking\_Present}$ = 2.64, $M_{TurnTaking\_Absent}$ = 0.53; $F(2, 222) = 15.522$, $p < .001$, $\eta^2 = .06$). and turn-initiation ($M_{TurnInitiation\_Present}$ = 2.47, $M_{TurnInitiation\_Absent}$ = 0.76; $F(2, 222) = 10.202$, $p < .01$, $\eta^2 = .04$) led to significantly more positive brand sentiment, while the interaction between the two factors was non-significant ($F(2, 222) = 0.212$, $p = .64$),

To further illustrate the pattern of results based on consumers expressed brand sentiment, Figure 5 shows the relationship between perceived humanness and brand sentiment across each experimental condition. Each observation is further color-coded based on the corresponding sentiment score (using the AFINN lexicon, see method section) and the most frequent sentiment for each data point (i.e., participant). These findings illustrate that greater perceived humanness is associated with a more positive brand sentiment and that the greatest density of positively valenced brand sentiment occurs under conditions of turn-taking and turn-initiation.

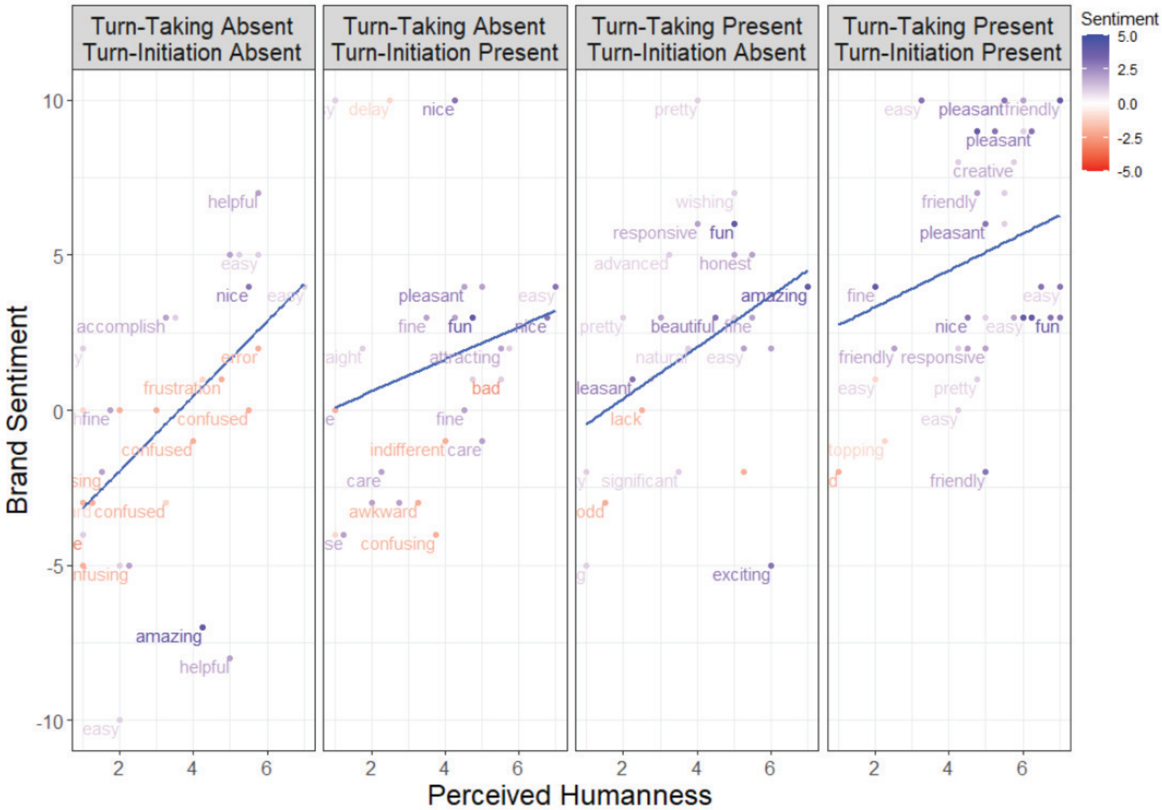



**Figure 2-5:** Brand sentiment as a function of perceived interface humanness

*Mediation Analysis.* We estimated a mediation model with 5000 bootstrap samples contrasting the interface with both turn-taking and turn-initiation versus all other conditions. Perceived interface humanness served as the mediator and brand sentiment as the dependent variable. This analysis revealed a significant indirect effect ($\beta_{Indirect} = .90$, 95% CI: [.31; 1.55]), indicating that the conversational properties of a brand's AI-based interface enhance brand sentiment via an increase in the perceived humanness of the interface.

*Discussion.* The findings of this study provide evidence of the critical role of turn-taking and turn-initiation in promoting greater perceptions of interface humanness and more positive brand sentiment. Specifically, the results suggest that consumers not only perceive the interface to possess more human characteristics under the presence (vs. absence) of turn-taking and turn-initiation, but that they also perceive the brand more positively, as reflected in a greater range of more positively-valenced brand sentiments (such as being fun, pleasant, and caring).

**Study 2**

Study 2 was designed with three key objectives in mind. First, while Study 1 used a qualitative measure to assess consumers' relationship with and perceptions of the brand (i.e., brand sentiment based on a thought-listing task), Study 2 provides a direct test of our hypothesis that turn-taking enhances consumers' experienced brand intimacy ($H_4$). Second, the current study was designed to further explore whether conversational interfaces also change consumer choice more directly, testing whether consumers accept a recommendation received during a shopping task. Third, we further assess whether the current effects are explained by a mere novelty effect, differences in ease-of-use perceptions, or task engagement.

*Design and Procedure.* All methods and statistical analyses for this study were pre-registered (https://aspredicted.org/blind.php?x=jf9ti7). A total of 391 consumers ($M_{Age} = 31.17$, $SD_{Age} = 11.51$; 64% females), recruited through Prolific Academic, completed the study. They were randomly assigned to a turn-taking present or turn-taking absent conversational interface condition (see Web Appendix A2.2 for screen shots). Participants' task was to specify a dress shirt as in the Pilot Study (using the same set of profiling questions, including questions about their body type, preferred shirt fit, eye and hair color, and style preferences, see Web Appendix A3.1). In the interface condition employing turn-taking, participants sequentially specified their answer taking one turn after the other whereas participants in the interface condition without turn-taking were asked all five profiling questions before indicating their preferred options. In both conditions the interface was able to initiate turns and participants were explicitly informed that they are interacting with an automated conversational interface.

After the profiling was completed, the interface recommended a shirt in a specific style and color to the participant, which they could either accept or reject and choose a different style from a set of available options (see Web Appendix 3.1). On completion of the shopping task, participants rated the perceived humanness of the system using the same four items as in Study 1 ($\alpha_{human}$ = .93, Web Appendix A5.1) and indicated how intimately they felt connected to the brand using three items (Thorbjørnsen et al., 2002: “I feel like SwissThread really cares about me”, “SwissThread really listens to what I have to say”, “I feel as though SwissThread really understands me”, 1 = “I do not agree at all” to 7 = “I fully agree”; $\alpha_{BrandIntimacy}$ = .90, see Web Appendix A5.2). Finally, to rule out that the current effects can be explained by a novelty effect or the interface being easier to use or greater task involvement, we assessed the perceived novelty when using the interface (Guiry, Mägi, and Lutz 2006; sample item: “The system offers novel experiences.”, $\alpha_{Novelty}$= .88), task involvement (Escalas, Moore, and Edell 2004; sample item: “The system really intrigued me.”, $\alpha_{Involvement}$= .90), and ease of use (Keller and Block 1997; sample item (reverse coded): “The system required a lot of effort.”, $\alpha_{Ease}$= .86). Details on all scale items can be found in Web Appendix A5.5-5.7.

*Manipulation Check.* At the end of the study, we assessed the effectiveness of the manipulation by asking participants to rate the perceived extent of turn-taking with the interface (adapted from Song and Zinkhan 2008; sample item: “The system facilitated a two-way conversation.”, $\alpha_{TurnTaking}$= .90, see Web Appendix A5.3). Supporting the effectiveness of the manipulation, the interface manipulation had a significant effect on the perception of turn-taking ($M_{TurnTaking\_Present}$ = 4.44, $M_{TurnTaking\_Absent}$ = 3.82; $F(1, 389) = 16.87$, $p < .001$, $\eta^2 = .04$).

*Results*

*Interface Humanness.* In line with our prediction ($H_1$) and replicating the results from Study 1, participants perceived a conversational interface with turn-taking as significantly more human-like compared to the interface without turn-taking ($M_{TurnTaking\_Present}$ = 4.59, $M_{TurnTaking\_Absent}$ = 3.73; $F(1, 389) = 30.62$, $p < .001$, $\eta^2 = .07$).

*Brand Intimacy.* Furthermore, we also find that consumers experienced a significantly more intimate relationship with the brand in the turn-taking condition compared to the no turn-taking condition ($M_{TurnTaking\_Present}$ = 4.52, $M_{TurnTaking\_Absent}$ = 3.94; $F(1, 389) = 13.62$, $p < .001$, $\eta^2 = .03$). Consistent with our theorizing ($H_4$), a mediation model with 5000 bootstrap samples confirmed that the effect on greater perceptions of brand intimacy was significantly mediated by the greater perception of humanness of the interface ($\beta_{Indirect} = .66$, 95% CI: [.42; .90]), rendering the residual direct effect non-significant ($p = .49$) and thus indicating full mediation.

*Recommendation Acceptance.* Next, we estimated a generalized linear model with a logistic link function to assess the effect of the experimental manipulation on consumers' decision to accept the recommended shirt option. Mirroring the findings on brand intimacy and perceptions of humanness, participants in the turn-taking present (vs. absent) condition were significantly more likely to accept the recommended shirt at the end of the shopping task ($P_{TurnTaking\_Present}$ = 57.6%, $P_{TurnTaking\_Absent}$ = 37.2%, $z = 4.01$, $p < .001$).

*Sequential Mediation.* As a more direct test of our theorizing, we estimated a sequential mediation model (5000 bootstrap samples). The interface manipulation was the independent variable, perceived humanness and brand intimacy were the proximal and distal sequential mediators, respectively, and acceptance of the recommended product was the dependent variable. The results (see Figure 6) show that, in support of our theorizing, the positive effect of turn-taking on recommendation acceptance was sequentially mediated by greater perceived interface humanness and increased brand intimacy ($\beta_{Indirect} = .17$, 95% CI: [.03; .35]). The two alternative, non-sequential mediational pathways were both non-significant with a CI including zero, while the residual direct effect of turn-taking on recommendation acceptance remained significant, indicating partial mediation.

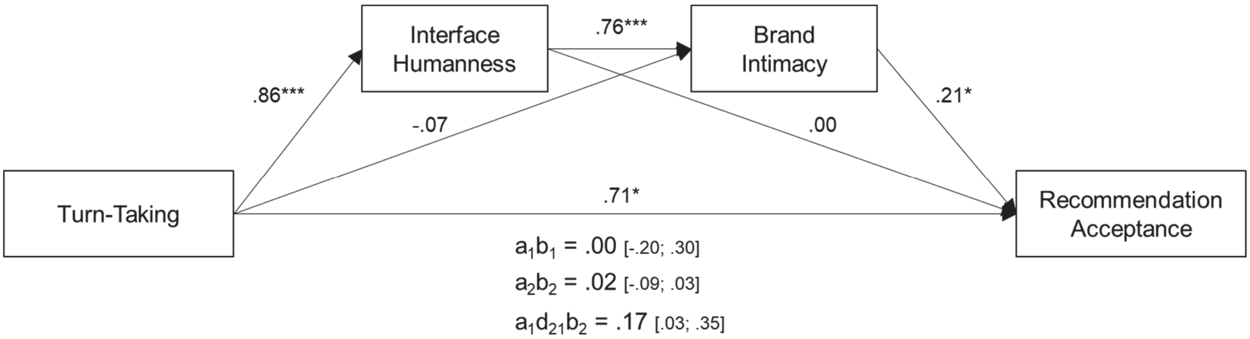



**Figure 2-6:** Sequential mediation model

The density plots shown in Figure 7 further illustrate that the probability mass of recommendation acceptance is shifted toward higher levels of perceived interface humanness and brand intimacy in the turn-taking present compared to the turn-taking absent conditions.

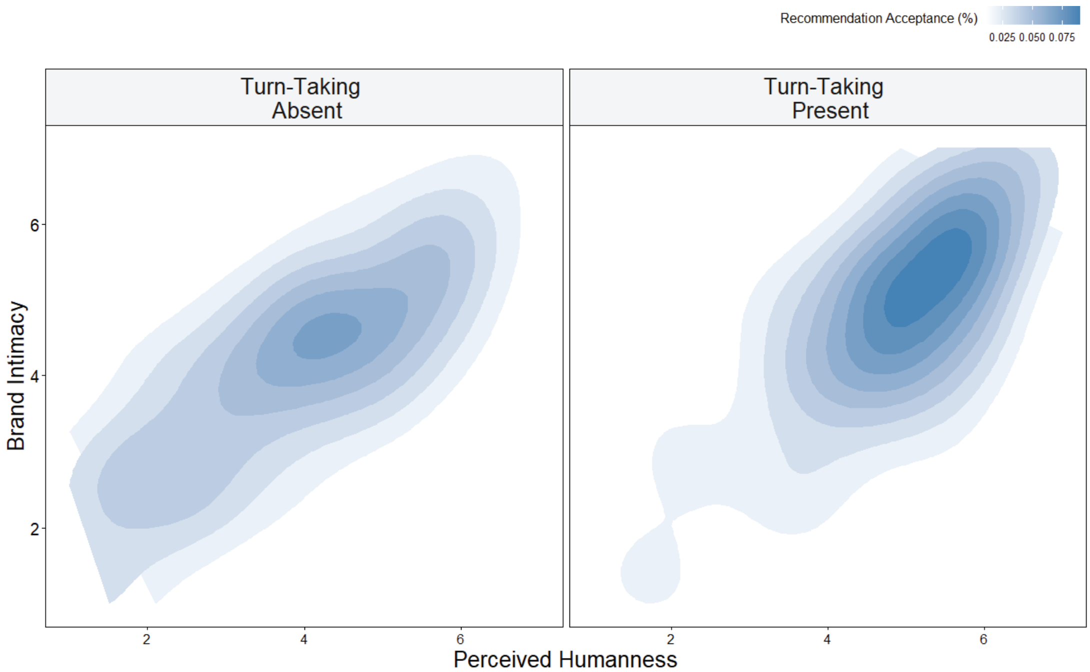



**Figure 2-7:** Recommendation acceptance as a function of interface humanness and brand intimacy

*Robustness Tests.* Next, we performed a series of robustness tests to further assess whether the effects of turn-taking on brand intimacy are explained by a mere novelty effect, differences in ease-of-use perceptions, or task engagement. We found that the turn-taking manipulation also had a significant impact on ease of use ($M_{TurnTaking\_Present}$ = 5.03, $M_{TurnTaking\_Absent}$ = 4.34; $F(1, 389) = 26.68$, $p < .001$, $\eta^2 = .06$), novelty ($M_{TurnTaking\_Present}$ = 4.76, $M_{TurnTaking\_Absent}$ = 4.38; $F(1, 389) = 6.11$, $p < .05$, $\eta^2 = .02$), and task involvement ($M_{TurnTaking\_Present}$ = 4.16, $M_{TurnTaking\_Absent}$ = 3.78; $F(1, 389) = 4.87$, $p < .05$, $\eta^2 = .02$). However, supporting the robustness of our theorizing, the effect of turn-taking on perceived interface humanness remained significant even after controlling for these three variables ($t(386) = 4.617$, $p < .001$). The same was true for the effect of interface humanness on brand intimacy ($t(386) = 14.624$, $p < .001$). We also assessed the robustness of the humanness mechanism using a series of Davidson-MacKinnon tests for non-nested models (Davidson and MacKinnon 1981). If ease of use, novelty, and involvement were sufficient to explain the increase in brand intimacy, then including the fitted values of the model testing humanness perceptions should provide no significant improvement in explanatory power. However, if adding the fitted values of humanness perception did improve prediction accuracy, we can conclude that the alternative mechanism model does not contain the correct set of regressors to predict brand intimacy (see also Greene 1993). While including the fitted values of ease of use, involvement, and novelty in the perceived humanness model revealed a significant improvement in model fit ($\theta_{HumannessModel}$ + Fitted(AlternativeModel), $t = 8.345$, $p < .001$), we found a comparatively stronger, significant improvement when accounting for the fitted values of the humanness mechanism in the alternative mechanism model ($\theta_{AlternativeModel}$ + Fitted(HumannessModel), $t = 14.624$, $p < .001$), thus supporting the importance of perceived interface humanness as a key mechanism to explain the increase in brand intimacy. Finally, we also controlled for the possibility that different consumer groups may be affected differently by the two interface conditions. In short, we found that all of the reported effects remained significant and were neither moderated by differences in participants' age (all $p$ values > .64) or gender (all $p$ values > .13), suggesting that the effect is robust across a broad range of consumer demographics.

*Discussion.* The results of Study 2 demonstrate the pivotal role of turn-taking in the ability of a brand's AI-based conversational interface to promote perceptions of humanness and foster intimate consumer-brand relationships, and influence consumer choice. In addition, they rule out potential alternative mechanisms via novelty perceptions, task involvement, or ease of use.

**Study 3**

One hallmark of human communication and effective conversations is the signaling of mutual understanding through conversational grounding. One objective of Study 3 was to examine whether semantic grounding affects perceptions of humanness and brand intimacy above and beyond turn-taking. In addition, this study was designed to provide further insight into the downstream consequences of interacting with a brand's AI-based conversational interface on brand-related behaviors, especially those related to future commitment to the brand (such as brand loyalty or consumers' willingness advocate the brand to other consumers).

*Design and Procedure.* All methods and statistical analyses for this study were pre-registered (https://aspredicted.org/blind.php?x=978y5t). A total of 585 participants ($M_{Age} = 31.95$, $SD_{Age} = 11.28$; 45% females), recruited through Prolific Academic, completed the study. They were randomly assigned to one of three interface conditions – baseline, turn-taking only, or turn-taking plus grounding (see Web Appendix A2.3 for screen shots). (An orthogonal manipulation of turn-taking and grounding in a full-factorial design was not possible since turn-taking is a prerequisite for grounding; Jurafsky and Martin 2019, see also Figure 1 for an illustration of the key conceptual properties). In all three conditions, participants were explicitly told that they are interacting with an automated conversational interface. They answered the same set of profiling questions as in Study 2, including questions about their body type, preferred shirt fit, eye and hair color, and style preferences (see Web Appendix A3.1 for details). In the interface conditions employing turn-taking, the interface asked one profiling question and only moved on to the next question after participants provided their answer, while in the interface condition without turn-taking, the system asked the same questions in the same order but all questions were presented simultaneously. Building on prior work on dialogue systems by Jurafsky and Martin (2017), we manipulated conversational grounding as follows. First, we used trivial acknowledgements signaling understanding after receiving a user input (such as "Got it, thanks!"). Second, we used paraphrasing and repeating back the user input to further strengthen the signal of the interface that it has received and recorded the user input. For example, if the user answered "slim" to the interfaces question of "Which shirt style do you prefer?", the interface would record the user input and repeat it back by saying "Good to know you prefer a slim shirt fit". These semantic properties did not alter turn-taking as they did not warrant a response from the user. Otherwise, the content and sequence of questions was identical across conditions and mirrored the experimental paradigm and procedures used in Study 2.

As in Study 2, we measured the perceived humanness of the interface ($\alpha_{human}$ = .93) and brand intimacy ($\alpha_{BrandIntimacy}$ = .90) using the same scales. Next, we measured consumers' willingness to pay a price premium based on four items (Jones, Taylor, and Bansal 2008; sample item: "I am likely to pay a little more for SwissThread"; $\alpha_{PayPremium}$ = .91) and willingness to advocate the brand to others using three items (Maxham and Netemeyer 2003; sample item: "I would recommend SwissThread as a shirt brand to my friends"; $\alpha_{RecommendIntent}$ = .96). Finally, we assessed participants' loyalty intention based on four items (adapted from Xiaoyun Han, Kwortnik, and Chunxiao Wang, 2008; sample item: "I'd like to continue using SwissThread in the future"; $\alpha_{PurchaseIntent}$ = .96). In addition to our primary dependent variables, we also measured consumers' brand rating using a single response item ("We will post a review of the customer experience people had with this customer service on a separate, crowd-sourced customer review platform. Please provide a star rating of the services below"; 5-point scale). (Details on all the scales can be found in Web Appendix A5.8-5.11.)

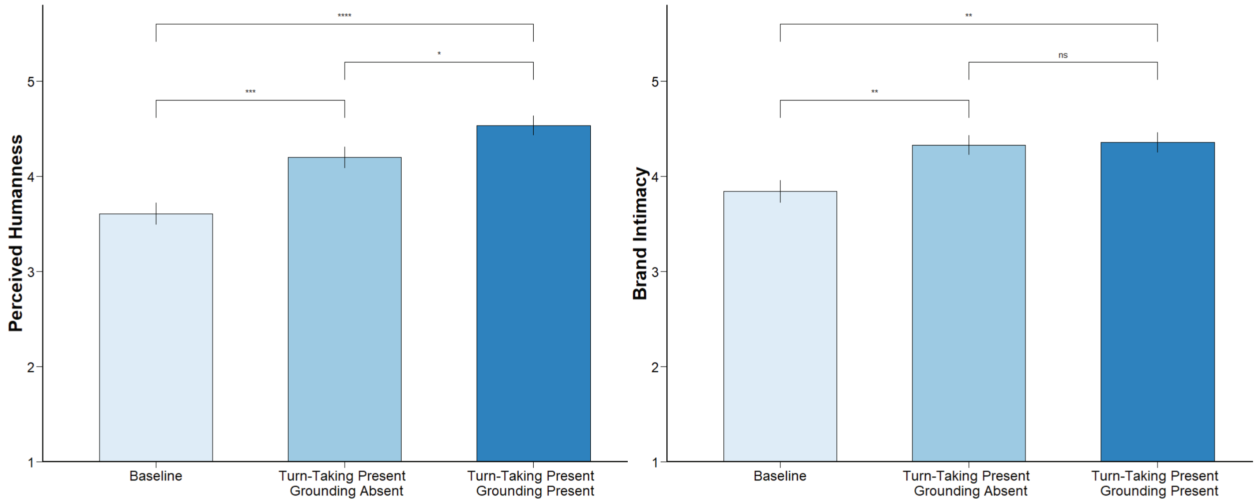
*Pretest*

To assess the effectiveness of the experimental manipulation, we conducted a pretest with 302 participants from the same population and using the same selection criteria as in the main study. Participants were randomly assigned to one of three conditions (same as in our main study). Specifically, they either interacted with a conversational interface with neither turn-taking nor grounding (baseline), a conversational interface with turn-taking but no grounding, or a conversational interface with both turn-taking and semantic grounding. In all three conditions, participants answered the same set of profiling questions as in the main study, including questions about their body type, preferred shirt fit, eye and hair color, and style preferences. Immediately after the interaction, we measured participants' perceived degree of grounding (adapted from Pierre and David 2006; sample item: "The system provided feedback of having perceived my input", $\alpha_{Grounding}$= .90, see Web Appendix A5.4). Supporting the effectiveness of our manipulation, the interface condition had a significant effect on the perception of grounding ($F(2, 299) = 35.07$, $p < .001$, $\eta^2 = .02$). Follow-up contrasts with Tukey HSD correction confirmed that the interface condition with grounding was perceived as significantly greater in grounding compared to both the turn-taking without grounding condition and relative to baseline ($M_{TurnTaking\_GroundingPresent} = 6.18$, $M_{Baseline} = 4.87$; $t(299) = 7.494$, $p < .001$; $M_{TurnTaking\_GroundingAbsent} = 5.03$; $t(299) = 6.890$, $p < .001$), while there was no difference in perceived grounding between turn-taking without grounding and baseline conditions ($p = .64$), supporting the effectiveness of the experimental manipulation.

*Results*

*Perceived Interface Humanness.* A one-way ANOVA with perceptions of humanness as the dependent variable revealed a significant overall effect of interface type ($F(2, 582) = 17.78$, $p < .001$, $\eta^2 = .06$; see Figure 8). In line with our theorizing ($H_3$), planned contrasts comparing the turn-taking with vs. without grounding conditions revealed a significant positive effect of grounding on the perceived humanness of the interface ($M_{TurnTaking\_GroundingPresent} = 4.53$, $M_{TurnTaking\_GroundingAbsent} = 4.20$, $t(582) = 2.165$, $p < .05$). Moreover, each of the two turn-taking conditions yielded greater perceived interface humanness than the baseline condition ($M_{Baseline} = 3.61$; both $p$ values $< .001$).

*Brand Intimacy.* The type of interface also had a significant overall effect on perceptions of brand intimacy ($F(2, 582) = 6.912$, $p < .01$, $\eta^2 = .02$). Planned contrasts revealed that brand intimacy was significantly higher in the turn-taking and grounding present condition than in the two other conditions ($t(1, 582) = 2.041$, $p < .05$). There was no difference in brand intimacy between the turn-taking with vs. without grounding conditions ($M_{TurnTaking\_GroundingPresent} = 4.36$, $M_{TurnTaking\_GroundingAbsent} = 4.33$, $p = .85$). However, as a conceptual replication of our key earlier finding, brand intimacy was significantly greater in the two turn-taking conditions compared to baseline ($M_{Baseline} = 3.84$, $t(582) = 3.714$, $p < .001$). These findings suggest that even though grounding enhances perceived interface humanness, it does not appear to further promote feelings of brand intimacy (above and beyond the effect of turn-taking).



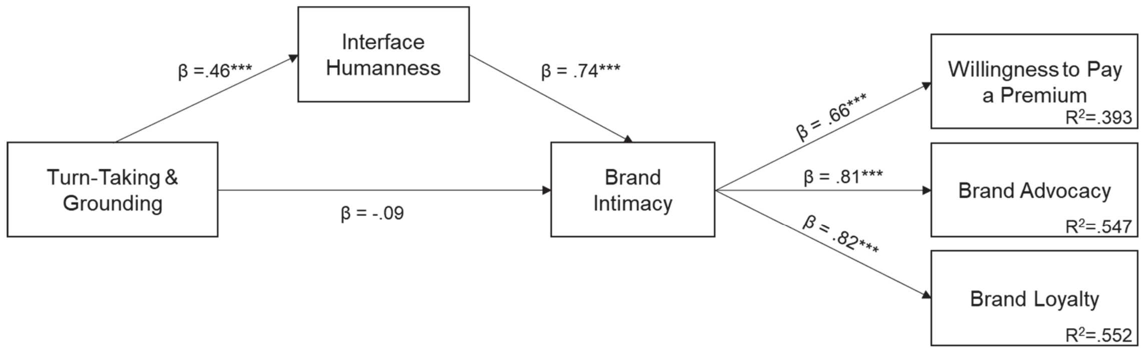

*Note:* *** = $p < .001$, ** = $p < .01$, * = $p < .05$;

**Figure 2-8:** Impact of turn-taking and grounding on interface humanness and brand intimacy

*Mediation Analysis.* We estimated a mediation model (5000 bootstrap samples) where the independent variable was turn-taking (binary coding: conditions with and without grounding combined vs. baseline), brand intimacy was the dependent variable, and perceived interface humanness was the mediator. Corroborating the results of Study 2, the positive effect of turn-taking (both with and without grounding) compared to baseline on brand intimacy ($\beta = .50$, $t = 3.714$, $p < .001$) was significantly mediated by enhanced perceptions of interface humanness ($\beta = .76$, $t = 5.589$, $p < .001$), which in turn led to an increase in brand intimacy ($\beta = .74$, $t = 27.117$, $p < .001$), rendering the residual direct effect of turn-taking on brand intimacy non-significant ($\beta = -.06$, $t = .656$, $p = .512$) and thus indicating full mediation ($\beta_{\text{Indirect}} = .56$, 95% CI: [.35; .77]). A Figure displaying this mediation model can be found in Web Appendix A7.1.

*Path Model.* We estimated a path model to further assess the impact of brand intimacy on behavioral brand outcomes ($H_5$). Specifically, we used the lavaan package in R with robust standard errors to model the effect of the interface condition (baseline coded as -1, turn-taking only coded as 0, and turn-taking plus grounding coded as 1) on humanness, how greater levels of humanness affect brand intimacy, and brand intimacy in turn drives subsequent brand-related behaviors. As summarized in Figure 9 and mirroring the main effect findings, a greater extent of turn-taking and grounding enhanced perceived humanness ($\beta = .463$, $z = 6.029$, $p < .001$), which in turn increased consumers' experienced brand intimacy ($\beta = .739$,

$z = 26.867$, $p < .001$). Most importantly, brand intimacy had a significant positive effect on consumers' willingness to pay a price premium ($\beta = .663$, $z = 19.867$, $p < .001$), future purchase intentions ($\beta = .815$, $z = 27.222$, $p < .001$), and brand advocacy ($\beta_{BrandIntimacy} = .809$, $z = 27.899$, $p < .001$). We also found that brand intimacy led to a significant boost in consumers' brand ratings ($\beta = .436$, $z = 20.062$, $p < .001$). These model results were robust even after controlling for differences in participants age and gender as indicated by higher information criteria compared to our main model ($BIC_{MainModel}$: 9953.6; $BIC_{ControlModel}$: 9932.1). Further details on contrast analyses for the behavioral brand outcomes can be found in Web Appendix A7.2.




*Note:* *** = $p < .001$, ** = $p < .01$, * = $p < .05$;

**Figure 2-9:** Path model

*Discussion.* The results of Study 3 demonstrate that grounding leads to greater perceived humanness of an interface, but that it does not necessarily further enhance the experience of brand intimacy above and beyond the effect of turn-taking. This provides a nuanced perspective on how conversational properties affect consumer-brand relationships, suggesting that turn-taking is the primary factor influencing both perceived interface humanness and brand intimacy, whereas the positive effect of grounding could be limited to making the conversational interface appear (even) more human and may not spill over to brand-specific outcomes. Study 3 further corroborates the pivotal (sequential) role of perceived interface humanness and brand intimacy in understanding how the properties of a brand's AI-based conversational interface affect consumers' overall evaluation of the brand and various brand-related behaviors.

## Study 4

The primary objective of this study was to test the hypothesis that the positive impact of the conversational properties of a brand's AI-based customer interface on consumers' perceptions of brand intimacy is amplified in settings that require higher levels of mutual understanding ($H_6$). In addition, we sought to shed light on how these conversational interface properties might affect consumers' preference for interacting with a brand's AI-based system in the future (instead of a human employee). Finally, this study was designed to demonstrate the robustness of the effects of the properties of conversational interfaces on brand intimacy by ruling out potential demand effects that might have resulted from measuring perceived interface humanness before brand intimacy.

*Design and Procedure.* All methods and statistical analyses for this study were pre-registered (https://aspredicted.org/blind.php?x=s7db9h). A total of 774 participants ($M_{Age}$ = 36.74, $SD_{Age}$ = 13.36; 56% females), recruited from Prolific Academic, completed the study. They were randomly assigned to a 2 (properties of the interface: baseline vs. turn-taking with grounding) × 2 (need for understanding: low vs. high) between-subjects design.

The following aspects were identical in all conditions. Participants used a conversational interface to order a to-go coffee with the brand Barista (see Web Appendix 2.4 for sample stimuli). They were informed that they would be interacting with an automated conversational interface. This interface did not include any other human elements (e.g., a human-like avatar). Participants' task was to order their preferred to-go coffee. They did so by answering a total of 10 questions to specify their order in terms of 10 different features (such as size, roast, type of milk, amount of sugar, etc.; see Web Appendix 3.3 for the full conversational flow). In all conditions, participants went through the same questions in exactly the same order to place their coffee order. Upon completion of the order process, the interface asked participants whether they would be willing to sign up for a loyalty program with Barista (as a behavioral brand-related outcome measure).

The properties of Barista's virtual customer interface were manipulated as follows. Participants placed their coffee order using either a baseline interface or one that included both turn-taking and grounding. In the baseline condition, the interface presented all 10 customization questions at the same time and asked the customer to specify their order by answering all of these questions in a single turn. By contrast, in the turn-taking with grounding condition, the interface presented one question at a time, waited for the customer to answer, and then acknowledged their answer before it proceeded to present the next question.

According to our theorizing, the effect of the conversational properties of a brand's interface on consumers' perceptions of brand intimacy is amplified in settings where the need for mutual understanding is greater ($H_6$). Naturally, this need increases as purchase decisions become more complex (see, e.g., Xu 2016). Thus, we manipulated the need for mutual understanding via the complexity of the choice task. In the high need for understanding condition, each of Barista's 10 coffee features had twice as many levels (i.e., options) as in the low need for understanding condition (see Dellaert, Donkers, and Van Soest 2012 for a similar approach to manipulating decision complexity). For example, there were either four sizes (small, medium, large, or extra-large) or only two sizes (small or large) to choose among in the high and low need for understanding conditions, respectively.

Following the interaction with Barista's interface, measures of perceived brand intimacy (as in all previous studies; $\alpha_{BrandIntimacy} = .87$) and self-brand connection (8-item scale adapted from Escalas 2004; $\alpha_{SelfBrandConnection} = .95$; see Web Appendix A5.12 for details) were obtained. Next, participants rated the perceived quality of Barista's products and service (2-item scale adapted from Hess, Ganesan, and Klein 2003; $\alpha_{Quality} = .92$; see Web Appendix A5.13 for details). Finally, we examined consumers' preference to interact with an AI-based conversational interface relative to a human employee. While prior work suggests that consumers might be reluctant to interact with AI (Longoni et al. 2019), we wanted to explore whether AI embedded in a brand's conversational interface can help overcome such reluctance by promoting more intimate consumer-brand relationships. We used a measure inspired by Longoni et al. (2019) – "Thinking about a future visit to the Barista coffee shop, please indicate who you would prefer to order your coffee with." (response options: human vs. virtual barista; see Web Appendix A5.14) – for this purpose.

*Pretest*

To assess the validity of the manipulation of need for understanding, we conducted a pretest with 150 participants from the same population. All participants were shown two coffee menus side-by-side, each with the same 10 customization features, but with one menu having twice as many feature levels than the other, exactly as in the main study. Participants were asked to indicate which of the two menus required a greater need for understanding ("Comparing the two menus, select the one that requires a greater need of being understood by the entity taking your order."). As intended, and confirming the effectiveness of the manipulation, the overwhelming majority of participants selected the more complex menu (94%, $\chi^2(1, 149) = 116$, $p < .001$).

*Results*

*Brand Intimacy.* Consistent with our theorizing, the interface employing turn-taking with grounding had a positive main effect on experienced brand intimacy relative to the baseline interface ($M_{TurnTaking\&Grounding} = 4.71$, $M_{Baseline} = 4.11$, $F(1, 772) = 32.16$, $p < .001$, $\eta^2 = .04$), whereas the main effect of need for understanding was not significant ($M_{HighNFU} = 4.48$, $M_{LowNFU} = 4.36$, $F(1, 772) = 1.09$, $p = .30$). Critically, in support of our directional hypothesis that the positive effect of the conversational interface properties on brand intimacy is amplified in settings that require higher levels of mutual understanding ($H_6$), we observed a significant interaction between the two factors ($F(1, 770) = 3.30$, $p < .05$, one-tailed, $\eta^2 = .004$). Although turn-taking and grounding had a significant positive impact on brand intimacy even when the need for understanding was low ($M_{TurnTaking\&Grounding} = 4.57$, $M_{Baseline} = 4.15$, $t(772) = 2.753$, $p < .01$), this effect was significantly more pronounced when the need for understanding was high ($M_{TurnTaking\&Grounding} = 4.86$, $M_{Baseline} = 4.06$, $t(772) = 5.280$, $p < .001$; see Figure 10). Finally, we found no effect of the experimental manipulation on the perceived quality of the coffee ($p < .29$).

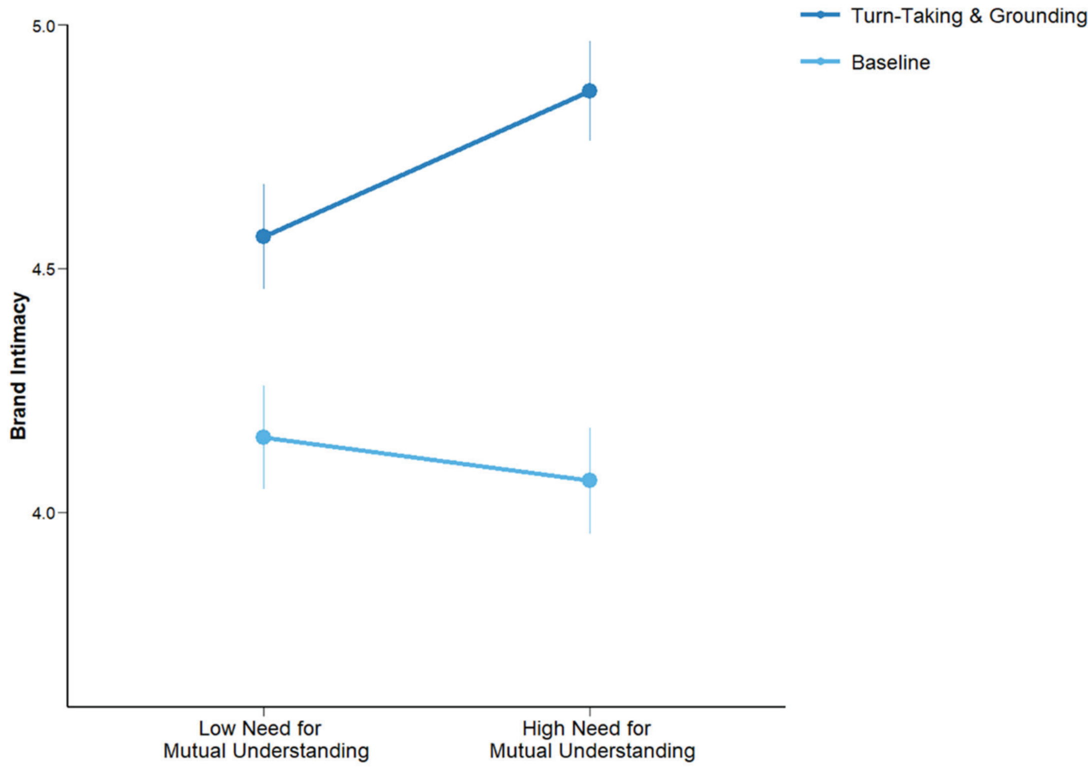



**Figure 2-10:** Effect of interface properties on brand intimacy: moderating role of need for mutual understanding

*Loyalty Card Sign-Up.* The interface properties of turn-taking and grounding had a positive main effect on willingness to sign-up to a loyalty program relative to the baseline interface ($M_{TurnTaking\&Grounding}$ = 77.1%, $M_{Baseline}$ = 64.1%, $F(1, 772) = 16.39$, $p < .001$, $\eta^2 = .02$), whereas the need for understanding had a negative effect on willingness to sign-up to a loyalty program ($M_{HighNFU}$ = 66.3%, $M_{LowNFU}$ = 75.2%, $F(1, 772) = 7.60$, $p < .01$, $\eta^2 = .01$). We also observed a significant interaction between the two factors ($F$(1, 772 = 3.56, $p < .05$, one-tailed, $\eta^2 = .005$). Consistent with our theorizing, the presence of turn-taking and grounding significantly increased sign-ups for Barista's loyalty card program when the need for mutual understanding was high ($M_{TurnTaking\&Grounding}$= 75.6%, $M_{Baseline}$ = 56.5%, $t(772) = 4.67$, $p < .01$), whereas this effect was not significant when the need for understanding was low ($p = .12$).

*Moderated Mediation.* We estimated a moderated mediation model with the conversational interface properties as the independent variable, loyalty card sign-up as the dependent variable, brand intimacy as the mediator, and need for understanding as a moderator of the effect of interface properties on brand intimacy. Consistent with the results reported above, the positive indirect effect of the conversational interface properties of turn-taking and grounding on loyalty card sign-up via enhanced brand intimacy is significant when need for understanding is low (β = .41, 95% CI: [.12; .71]), and this effect is significantly amplified when need for understanding is high (β = .80, 95% CI: [.50; 1.10]), as indicated by a significant moderated mediation index (β = .10, 95% CI: [.00; .23]).

*Self-Brand-Connection.* The conversational interface properties had a positive main effect on self-brand connection ($M_{TurnTaking\&Grounding}$= 3.16, $M_{Baseline}$ = 2.84, $F(1, 772) = 9.01$, $p < .01$, $\eta^2 = .01$). The main effect of need for understanding was marginally significant ($M_{HighNFU}$ = 3.11, $M_{LowNFU}$ = 2.91, $F(1, 772) = 3.47$, $p = .063$, $\eta^2 = .004$) and the interaction between the two factors was not significant ($p = .50$). Critically, the effect of the interface properties on brand intimacy remained significant even after controlling for self-brand connection ($F(1, 772) = 4.90$, $p < .05$). As expected, the measure of self-brand connection was positively correlated with brand intimacy (r = .695, $t(772) = 26.8$, $p < .001$).

*Human vs. AI-Based System Preference.* A generalized linear model with a logistic link function revealed a significant positive main effect of the conversational interface properties (of turn-taking and grounding) on the proportion of participants who indicated that they would prefer to place a future coffee order at Barista with the virtual customer interface instead of a human employee ($P_{TurnTaking\&Grounding}$ = 45.7%, $P_{Baseline}$ = 35.9%, $z = 2.35$, $p < .05$). Neither the main effect of need for understanding ($p = .26$) nor the interaction among the two factors ($p = .58$) were significant.

*Discussion.* Study 4 provides evidence in support of the hypothesized moderating role of the extent to which a setting requires mutual understanding between a consumer and a brand ($H_6$). In particular, it demonstrates that the brand-intimacy-enhancing effect of the brand's conversational interface is pronounced in settings that require greater mutual understanding. In addition, the results of Study 4 show that "humanizing" a brand's AI-based customer interface by incorporating essential conversational properties can promote consumers' acceptance of brands' use of other AI technologies in the future.

## General Discussion

Across four studies, we shed light on how fundamental properties of human dialogue (turn-taking, turn-initiation, and grounding) shape how consumers perceive and interact with a brand's AI-based conversational interface, inducing greater perceptions of humanness, leading to more intimate consumer-brand relationships, and ultimately affecting a series of downstream consequences for consumers (both in terms of the immediate persuasiveness of the brand as well as a stronger future commitment to the brand). We find evidence for these effects across multiple product settings (from design and fashion contexts to automating food or beverage delivery). Specifically, Study 1 shows that the ability to take turns and the ability to initiate turns results in greater levels of perceived humanness and more positive brand sentiment. Study 2 provides more direct evidence that turn-taking and turn-initiation promote brand intimacy and make consumers more willing to accept a recommendation for a target product. Study 2 also revealed that the use of conversational interfaces renders the same task more involving and novel, highlighting that the underlying psychological mechanism is arguably multiply determined. Study 3 extends these findings to other downstream consequences (willingness to pay a price premium, willingness to advocate the brand to others and future purchase intentions from the brand), while Study 4 provides evidence for the underlying psychological process of the conversational properties to establish a sense of mutual understanding, showing that a conversational interface's ability to promote a more intimate relationship with the brand is enhanced in situations that require higher (as opposed to lower) levels of mutual understanding.

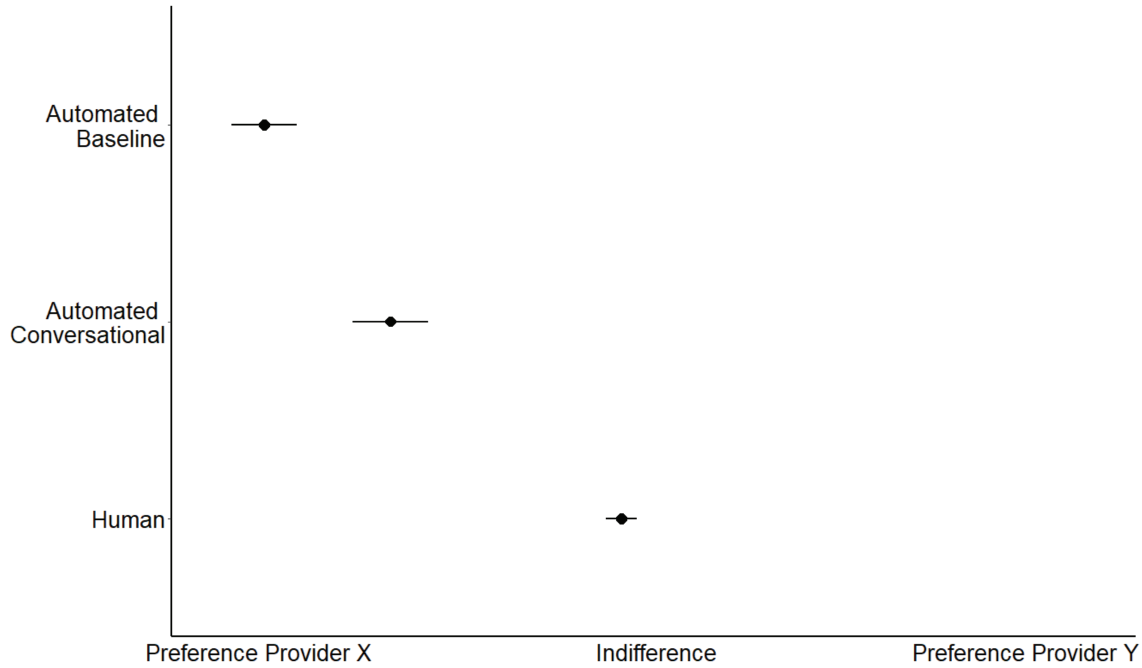
*Theoretical Implications.* The current findings make five novel theoretical contributions. First, the current work is the first line of systematic research showing that conversational interfaces can act as a malleable technology to alter consumers' brand experience, proposing a novel conceptual model of verbal embodiment through branded conversational interfaces. These findings link previously disconnected streams of research on humanizing brands and brand interactions (Holzwarth et al. 2006; MacInnis and Folkes 2017; Schroll, Schnurr, and Grewal 2018; Verhagen et al. 2014; Wang et al. 2007) with research on embodied systems (Bickmore and Cassell 2001a; Cassell et al. 2000, 2001), and we identify a series of critical design characteristics in terms of structural (turn-taking, turn-initiation) and semantic (grounding) interface properties that facilitate the humanization of consumer-brand interactions. Our findings also demonstrate the underlying process of how these properties establish a sense of understanding with a branded interface, a critical aspect of designing more relatable AI (Sciutti et al. 2018).

Second, our findings contribute to recent research on how embodied actions can facilitate emotional connections with a brand (Hadi and Valenzuela 2014), suggesting that a conversational interface acts as a new form of embodied consumer-brand interaction, influencing brand intimacy, and a broad range of behavioral brand outcomes (recommendation acceptance, willingness to pay a price premium, brand loyalty, and brand advocacy). The current work contributes to recent calls for research on technology-mediated brand interactions and consumer behavior (Melumad et al. 2020; Yadav and Pavlou 2014) providing a novel perspective on the impact of branded conversational interfaces to shape, influence and ultimately "design" consumer-brand relationships. This also highlights that conversational interfaces are more than just tools for service automation and should be considered as an integral part of a brand's technology-mediated branding strategy.

Third, the current research also contributes to recent work on how human consumers engage with inanimate objects and how conversational interfaces fundamentally change and eventually create a new "assemblage" of consumer-brand experiences (Hoffman and Novak 2018; Novak and Hoffman 2018). Specifically, the current work provides evidence that even though a conversational interface acts merely as a medium to complete a task (such as ordering a cup of coffee) consumers tend to form a relationship with the medium and develop (or effectively attribute) meaning toward the assemblage, which extends toward the brand embodied through the medium (i.e., the chatbot). Thus, the current work contributes to recent work by Hoffman and Novak (2018), providing initial evidence on how consumer-object assemblages alter consumer shopping experiences by creating a more intimate consumer-brand relationship.

Fourth, our findings also contribute to the literature on persuasion effects in branding and marketing. The majority of prior work explored the effects of how human actors representing a brand (such as sales representatives or service employees) can promote compliance during a service delivery or shopping episode. These influence tactics have become increasingly less effective due to consumer learning (Friestad and Wright 1994), "market meta-cognition" (Wright 2002) and a general skepticism with respect to offerings received by human service providers for which consumers have developed various response strategies (Kirmani and Campbell 2004). The current work demonstrates that consumers are substantially more susceptible to such influence attempts delivered through conversational interfaces that mimic principles of human communication (turn-taking, turn-initiation, and grounding) to appear progressively more human, affecting consumers' likelihood to pay a price premium, greater recommendation acceptance, and recommending the brand to others. The comparatively weaker effects that we observed for grounding also open up the door for future research on other grounding mechanisms that go beyond verbal forms of grounding and to assess whether additional visual forms of grounding (such as accompanying messages from an animated avatar with gestural forms of grounding, like head nodding) can further enhance the persuasiveness of AI-based conversational interfaces.

Finally, the findings of this research also make a broader contribution to the emerging literature on consumer resistance to AI, and the tendency of consumers to view AI as cold and mechanical (Dietvorst et al. 2018; Granulo et al. 2019; Longoni et al. 2019). In their recent work on resistance to medical AI, Longoni et al. (2019) suggest that the lack of warmth and empathy may explain consumers' resistance to trust and to choose AI over a human to complete a (medical screening) task. In order to assess whether conversational interfaces may counter consumers' resistance to AI through the intimacy-enhancing effect revealed in this research, we conducted a pilot study building on the paradigm used in Longoni et al. (2019; Study 3A). Specifically, participants (N = 258; $M_{Age} = 34.53$, $SD_{Age} = 12.69$; 62% female) read a description about a skin cancer screening and were randomly assigned to one of three conditions. Participants had to choose their preferred treatment from either (1) two human dermatologists, (2) a human dermatologist or an automated dermatologist (described as an algorithm assessing a patient's medical condition), or (3) a human dermatologist or a conversational dermatologist interface (see Web Appendix A8 for the full details on stimuli and experimental set-up). The first two conditions resembled the original conditions in Longoni et al. (2019), whereas the conversational interface condition employed an interface using turn-taking, turn-initiation, and grounding throughout the screening task and consumers had a short interaction with the interface inspired by the original description in Longoni et al. (2019). A one-way ANOVA assessing participants' preferred medical provider revealed a significant effect of the provider condition ($F(2, 255) = 38.33$, $p < .001$, $\eta^2 = .23$). Follow-up contrasts with Tukey HSD correction revealed that, in line with Longoni et al.'s (2019) finding, consumers generally preferred a human over an automated provider ($M_{Human} = 3.85$, $M_{AutomatedBaseline} = 2.22$, $t = 5.439$, $p < .001$). However, as illustrated in Figure 11, consumer preference shifted significantly toward a conversational interface employing turn-taking, turn-initiation, and grounding relative to the baseline of an automated medical provider ($M_{AutomatedConversational}$ =2.80, $t$ = 2.969, $p$ < .01), while still significantly lower compared to the human provider condition ($t$ = 5.439, $p$ < .001). Although these findings can only provide initial, preliminary evidence, they suggest that developing more relatable technology through a greater extent of turn-taking, turn-initiation, and grounding holds the potential to significantly decrease consumers' resistance toward AI.




*Note*: Preference for Provider X always resembled a human dermatologist. Preference for Provider Y resembled, depending on the experimental condition, consumers' preference for either a human dermatologist, automated dermatologist, or an automated conversational interface.

**Figure 2-11:** Preferences of human vs. automated medical provider

*Practical Implications.* To the best of our knowledge, the current work is the first to provide a theory-driven framework to systematically develop more relatable and effective AI-based conversational interfaces. Specifically, we provide a theory-driven framework to enhance consumer-brand relationships by utilizing fundamental principles of human-to-human communication (turn-taking, turn-initiation and grounding). Current industry practice is largely agnostic with respect to these structural (turn-taking, turn-initiation) and semantic (grounding) properties employed by a brand's conversational interface. As highlighted in the current research, changing any of these dimensions significantly affects consumers' perception of the interface, shapes consumer-brand relationships, and systematically impacts consumer choice (from recommendation acceptance to loyalty decisions). Thus, the current work offers important insights and design guidelines for practitioners on how to better and more rigorously engineer a brand's conversational interface.

Finally, a major issue in modern brand management is to harmonize consumer brand perceptions across touch-points and interaction episodes. The current work highlights a novel opportunity by employing conversational interfaces to "harmonize" this experience through consumer-brand touch-points. Contrary to current industry practice of training service employees in both online and offline channels, conversational interfaces are likely to become a cost-effective strategy in marketers' brand management toolbox, providing both a high level of control as well as consistency in the way how brands engage with their customers.

*Conclusion.* Conversational interfaces are more than cost-effective interactive tools to automate consumer-firm interactions. The current work shows that AI-based conversational interfaces provide a powerful branding ecosystem to deliver highly scalable, one-to-one consumer-brand experiences, nurture consumer-brand relationships, and systematically influence a broad range of behavioral brand outcomes (from greater recommendation acceptance to increased brand loyalty). The theory-driven framework introduced and tested here advances our understanding of the important role that AI-based conversational interfaces might play in the formation and deepening of consumer-brand relationships.

## References


Araujo, Theo (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, 183–89, https://doi.org/10.1016/j.chb.2018.03.051.

Baumeister, Roy F. and Mark R. Leary (1995), "The Need to Belong: Desire for Interpersonal Attachments as a Fundamental Human Motivation," *Psychological Bulletin*, 117(3), 497–529.

Bellur, Saraswathi and S Shyam Sundar (2017), "Talking Health With a Machine : How Does Message Interactivity Affect Attitudes and Cognitions ?," *Human Communication Research*, 43, 25–53.

Bertacchini, Francesca, Eleonora Bilotta, and Pietro Pantano (2017), "Shopping with a Robotic Companion," *Computers in Human Behavior*, 77, 382–95, https://doi.org/10.1016/j.chb.2017.02.064.

BI Intelligence (2016), "80% of Businesses Want Chatbots by 2020," *Business Insider*, https://www.businessinsider.com/80-of-businesses-want-chatbots-by-2020-2016-12?r=US&IR=T.

Bickmore, Timothy and Justine Cassell (2001), "Relational Agents: A Model and Implementation of Building User Trust," *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2(2), 396–403.

Brick, Danielle J. and Susan Fournier (2017), "How Consumers' Attitudes Toward Brands Are Shaped," in *The Routledge Companion to Consumer Behavior*, ed. Michael R. Solomon and Tina M. Lowrey, Abingdon: Routledge, 227–42, https://doi.org/10.4324/9781315526935.

Brown, Lew G. (1989), "The Strategic and Tactical Implications of Convenience in Consumer Product Marketing," *Journal of Consumer Marketing*, 6(3), 13–19.

Carroll, Barbara A. and Aaron C. Ahuvia (2006), "Some Antecedents and Outcomes of Brand Love," *Marketing Letters*, 17(2), 79–89.

Cassell, Justine, Timothy Bickmore, L. Campbell, H. Vilhjálmsson, and H. Yan (2001), "More Than Just a Pretty Face: Conversational Protocols and the Affordances of Embodiment," *Knowledge-Based Systems*, 14(1–2), 55–64.

Cassell, Justine, Joseph Sullivan, Scott Prevost, and Elizabeth Churchill (2000), *Embodied Conversational Agents*, ed. Justine Cassell, Joseph Sullivan, Scott Prevost, and Elizabeth Churchill, Cambridge, MA: MIT Press.

Cassell, Justine and Kristinn R. Thórisson (1999), "The Power of a Nod and a Glance: Envelope vs. Emotional Feedback in Animated Conversational Agents," *Applied Artificial Intelligence*, 13(4–5), 519–38.

Cialdini, Robert B and B. J. Sagarin (1994), "Principles of Interpersonal Influence," in *Persuasion: Psychological Insights and Perspectives*, ed. Timothy C. Brock and Melanie C. Green, Thousand Oaks, CA: Sage Publications, Inc., 143–69.

Clark, Herbert H. and Susan E. Brennan (1991), "Grounding in Communication," in *Perspectives on Socially Shared Cognition*, 222–33.

Clark, Herbert H. and Edward F. Schaefer (1989), "Contributing to Discourse," *Cognitive Science*, 13(2), 259–94.

Clark, Jon (2017), "Everything You Need To Know About Chatbot Marketing Forbes Agency

Council," *Forbes Magazine*, https://www.forbes.com/sites/forbesagencycouncil/2017/12/08/everything-you-need-to-know-about-chatbot-marketing/#79a85500509a.

Comendador, Benilda E. V., Bien M. B. Francisco, Jefferson S. Medenilla, Sharleen M. T. Nacion, and Timothy B. E. Serac (2015), "Pharmabot: A Pediatric Generic Medicine Consultant Chatbot," *Journal of Automation and Control Engineering*, 3(2), 137–40, http://www.joace.org/index.php?m=content&c=index&a=show&catid=42&id=218.

Dale, Robert (2016), "The Return of the Chatbots," *Natural Language Engineering*, 22(5), 811–17.

Davidson, Russell and James G. MacKinnon (1981), "Several Tests for Model Specification in the Presence of Alternative Hypotheses," *Econometrica*, 49(3), 781–93.

Deci, Edward L. and Richard Flaste (1995), *Why We Do What We Do: The Dynamics of Personal Autonomy*, New York, NY, US: G P Putnam's Sons.

Deci, Edward L and Richard M Ryan (2000), "The " What " and " Why " of Goal Pursuits: Human Needs and the Self-Determination of Behavior," *Psychological Inquiry*, 11(4), 227–68.

Dellaert, Benedict G.C., Bas Donkers, and Arthur Van Soest (2012), "Complexity Effects in Choice Experiment-Based Models," *Journal of Marketing Research*, 49(3), 424–34.

Dellaert, Benedict G C and Stefan Stremersch (2005), "Marketing Mass-Customized Products: Striking a Balance Between Utility and Complexity," *Journal of Marketing Research*, XLII(May), 219–27.

Dietvorst, Berkeley J., Joseph P. Simmons, and Cade Massey (2018), "Overcoming Algorithm Aversion: People Will Use Imperfect Algorithms If They Can (Even Slightly) Modify Them," *Management Science*, 64(3), 1155–70.

Erdem, Tülin and Joffre Swait (1998), "Brand Equity as a Signaling Phenomenon," *Journal of Consumer Psychology*, 7(2), 131–57.

Escalas, Jennifer Edson (2004), "Narrative Processing: Building Consumer Connections to Brands," *Journal of Consumer Psychology*, 14(1–2), 168–80.

Escalas, Jennifer Edson, Marian Chapman Moore, and Julie A. Edell (2004), "Fishing for Feelings? A Hook Helps!," *Journal of Consumer Psychology*, 14(1), 105–13, http://www.sciencedirect.com/science/article/pii/S1057740804701370.

Finch, John F. and Robert B Cialdini (1989), "Another Indirect Tactic of (Self-) Image Managment: Boosting," *Personality and Social Psychology Bulletin*, 15(2), 222–32.

Fournier, Susan (1998), "Consumers and Their Brands: Developing Relationship Theory in Consumer Research," *Journal of Consumer Research*, 24(4), 343–53, https://academic.oup.com/jcr/article-lookup/doi/10.1086/209515.

Friestad, Marian and Peter Wright (1994), "The Persuasion Knowledge Model: How People Cope with Persuasion Attempts," *Journal of Consumer Research*, 21, 1–30, https://academic.oup.com/jcr/article-lookup/doi/10.1086/209380.

Gartner (2018), "Gartner Says 25 Percent of Customer Service Operations Will Use Virtual Customer Assistants by 2020," *Gartner*, https://www.gartner.com/en/newsroom/press-releases/2018-02-19-gartner-says-25-percent-of-customer-service-operations-will-use-virtual-customer-assistants-by-2020.

Ghose, Supratip and Jagat Joyti Barua (2013), "Toward the Implementation of a Topic Specific Dialogue Based Natural Language Chatbot as an Undergraduate Advisor," *2013 International Conference on Informatics, Electronics and Vision, ICIEV 2013*, 1–5.

Gilbert, Shirley J (1976), "Self Disclosure, Intimacy, and Communication in Families," *The Family Coordinator*, 25(3), 221–31.

Go, Eun and S. Shyam Sundar (2019), "Humanizing Chatbots: The Effects of Visual, Identity and Conversational Cues on Humanness Perceptions," *Computers in Human Behavior*, 97.

Goldstein, Noah J., Robert B. Cialdini, and Vladas Griskevicius (2008), "A Room with a Viewpoint: Using Social Norms to Motivate Environmental Conservation in Hotels," *Journal of Consumer Research*, 35(3), 472–82, https://academic.oup.com/jcr/article-lookup/doi/10.1086/586910.

Gong, Li (2008), "How Social Is Social Responses to Computers ? The Function of the Degree of Anthropomorphism in Computer Representations," *Computers in Human Behavior*, 24, 1494–1509.

Granulo, Armin, Christoph Fuchs, and Stefano Puntoni (2019), "Psychological Reactions to Human versus Robotic Job Replacement," *Nature Human Behaviour*, 3(10), 1062–69, http://dx.doi.org/10.1038/s41562-019-0670-y.

Greene, William H. (1993), *Econometric Analysis*, 2nd ed., New York, NY, US: Macmillan Publishing Company.

Grice, H.P. (1975), *Logic and Conversation*, 3 Syntax and Semantics, Volume 3: Speech Acts, ed. Peter Cole and Jerry L. Morgan, Academic Press.

Grice, Paul (1989), "Logic and Conversation," in *Studies in the Way of Words*, Cambridge, MA: Harvard University Press, 22–40.

La Guardia, Jennifer G., Richard M. Ryan, Charles E. Couchman, and Edward L. Deci (2000), "Within-Person Variation in Security of Attachment: A Self-Determination Theory Perspective on Attachment, Need Fulfillment, and Well-Being," *Journal of Personality and Social Psychology*, 79(3), 367–84.

Guiry, Michael, Anne W. Mägi, and Richard J. Lutz (2006), "Defining and Measuring Recreational Shopper Identity," *Journal of the Academy of Marketing Science*, 34(1), 74–83.

Hadi, Rhonda and Ana Valenzuela (2014), "A Meaningful Embrace: Contingent Effects of Embodied Cues of Affection," *Journal of Consumer Psychology*, 24(4), 520–32, http://dx.doi.org/10.1016/j.jcps.2014.02.001.

Heikes, Brian (2017), "Is Your Brand Developing a Chatbot Strategy?," *Retaildive*, https://www.retaildive.com/ex/mobilecommercedaily/is-your-brand-developing-a-chatbot-strategy.

Hess, Ronald L. Jr., Shankar Ganesan, and Noreen M Klein (2003), "Service Failure and Recovery : The Impact of Relationship Factors on Customer Satisfaction," *Journal of the Academy of Marketing Science*, 31, 127–45.

Hildebrand, Christian and Anouk Bergner (2020), "Conversational Robo Advisors as Surrogates of Trust : Onboarding Experience , Firm Perception , and Consumer Financial Decision Making," *Journal of Academy of Marketing Science*.

Hildebrand, Christian, Gerald Häubl, and Andreas Herrmann (2014), "Product Customization via Starting Solutions," *Journal of Marketing Research*, LI(December), 707–25.

Hirschberg, Julia and Christopher D Manning (2015), "Advances in Natural Language Processing," *Science*, 349(6245), 261–66.

Hoffman, Donna L. and Thomas P. Novak (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.

Holzwarth, Martin, Chris Janiszewski, and Marcus M Neumann (2006), “The Influence of Avatars on Online Consumer Shopping Behavior,” *Journal of Marketing*, 70(4), 19–36, http://journals.ama.org/doi/abs/10.1509/jmkg.70.4.19.

Jin, Seung A.Annie and Yongjun Sung (2010), “The Roles of Spokes-Avatars’ Personalities in Brand Communication in 3D Virtual Environments,” *Journal of Brand Management*, 17(5), 317–27.

Johnston Taylor, Susan (2016), “Very Human Lessons From Three Brands That Use Chatbots To Talk To Customers,” *The Fast Company*, https://www.fastcompany.com/3064845/human-lessons-from-brands-using-chatbots.

Jones, Tim, Shirley F. Taylor, and Harvir S. Bansal (2008), “Commitment to a Friend, a Service Provider, or a Service Company-Are They Distinctions Worth Making?,” *Journal of the Academy of Marketing Science*, 36(4), 473–87.

Jurafsky, Daniel and James Martin (2017), “Dialog Systems and Chatbots,” in *Speech and Language Processing*, Pearson, 418–40.

——— (2019), “Dialogue Systems and Chatbots,” in *Speech and Language Processing*, Draft of October 2, 2019.

Keller, Punam Anand and Lauren G. Block (1997), “Vividness Effects: A Resource-Matching Perspective,” *Journal of Consumer Research*, 24(3), 295–304.

Kirmani, Amna and Margaret C. Campbell (2004), “Goal Seeker and Persuasion Sentry: How Consumer Targets Respond to Interpersonal Marketing Persuasion,” *Journal of Consumer Research*, 31(3), 573–82, https://academic.oup.com/jcr/article-lookup/doi/10.1086/425092.

Landis, Geoffrey A. (2014), “Future Tense: The Chatbot and the Drone,” *Communications of the ACM*, 57(7), 111–12.

Lee, Sangwon, Naeun Lee, and Young June Sah (2019), “Perceiving a Mind in a Chatbot: Effect of Mind Perception and Social Cues on Co-Presence, Closeness, and Intention to Use,” *International Journal of Human-Computer Interaction*, 7318, https://doi.org/10.1080/10447318.2019.1699748.

Levinson, Stephen C. (2016), “Turn-Taking in Human Communication - Origins and Implications for Language Processing,” *Trends in Cognitive Sciences*, 20(1), 6–14.

Li, Xueni, Kimmy Wa Chan, and Sara Kim (2019), “Service with Emoticons: How Customers Interpret Employee Use of Emoticons in Online Service Encounters,” *Journal of Consumer Research*, 45(5), 973–87.

Lokman, Abbas Saliimi and Jasni Mohamad Zain (2010), “Extension and Prerequisite: An Algorithm to Enable Relations between Responses in Chatbot Technology,” *Journal of Computer Science*, 6(10), 1212–18.

Longoni, Chiara, Andrea Bonezzi, and Carey K Morewedge (2019), “Resistance to Medical Artificial Intelligence,” *Journal of Consumer Research*, 46, 629–50.

MacInnis, Deborah J. and Valerie S. Folkes (2017), “Humanizing Brands: When Brands Seem to Be like Me, Part of Me, and in a Relationship with Me,” *Journal of Consumer Psychology*, 27(3), 355–74, http://dx.doi.org/10.1016/j.jcps.2016.12.003.

Maner, Jon K., Carol L. Luce, Steven L. Neuberg, Robert B. Cialdini, Stephanie Brown, and Brad J. Sagarin (2002), “The Effects of Perspective Taking on Motivations for Helping: Still No Evidence for Altruism,” *Personality and Social Psychology Bulletin*, 28(11), 1601–10.

Maxham, James G. and Richard G. Netemeyer (2003), “Firms Reap What They Sow: The Effects of Shared Values and Perceived Organizational Justice on Customers’ Evaluations of Complaint Handling,” *Journal of Marketing*, 67(1), 46–62.

McAdams, Dan P., R. Jeffrey Jackson, and Carol Kirshnit (1984), “Looking, Laughing, and Smiling in Dyads as a Function of Intimacy Motivation and Reciprocity,” *Journal of Personality*, 52(3), 261–73.

McInnes, Fergus and David Attwater (2004), “Turn-Taking and Grounding in Spoken Telephone Number Transfers,” *Speech Communication*, 43(3), 205–23.

Melumad, Shiri, Rhonda Hadi, Christian Hildebrand, and Adrian F. Ward (2020), “Technology-Augmented Choice: How Digital Innovations Are Transforming Consumer Decision Processes,” *Customer Needs and Solutions*.

Metz, Cade (2020), “Riding Out Quarantine With a Chatbot Friend: ‘I Feel Very Connected,’” *The New York Times*, https://www.nytimes.com/2020/06/16/technology/chatbots-quarantine-coronavirus.html.

Nagao, Katashi and Akikazu Takeuchi (1994), “Speech Dialogue with Facial Display: Multimodal Human-Computer Conversation,” in *CHI ’94 Conference Companion on Human Factors in Computing Systems*, Bostob, MA, USA, 449–50.

Natarelli, Mario and Rina Plapler (2017), *Brand Intimacy: A New Paradigm in Marketing*, United States: Hatherleigh Press.

Nielsen, Finn Årup (2011), “A New ANEW: Evaluation of a Word List for Sentiment Analysis in Microblogs,” *CEUR Workshop Proceedings*, 718, 93–98.

Noussair, Charles, Stephane Robin, and Bernard Ruffieux (2004), “Revealing Consumers’ Willingness-to-Pay: A Comparison of the BDM Mechanism and the Vickrey Auction,” *Journal of Economic Psychology*, 25(6), 725–41.

Novak, Thomas and Donna L. Hoffman (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*, 216–37.

op den Akker, Rieks and Merijn Bruijnes (2012), “Computational Models of Social and Emotional Turn-Taking for Embodied Conversational Agents: A Review,” *Technical Report CTIT*, 1–62.

Park, C. Whan, Deborah J. Maclnnis, Joseph Priester, Andreas B. Eisingerich, and Dawn Lacobucci (2010), “Brand Attachment and Brand Attitude Strength: Conceptual and Empirical Differentiation of Two Critical Brand Equity Drivers,” *Journal of Marketing*, 74(6), 1–17.

Pierre, D. and T. David (2006), “Sharing Solutions: Persistence and Grounding in Multimodal Collaborative Problem Solving,” *Journal of the Learning Sciences*, 15(1), 10–34, http://en.scientificcommons.org/23809736.

Ramsey, Rosemary P. and Ravipreet S. Sohi (1997), “Listening to Your Customers: The Impact of Perceived Salesperson Listening Behavior on Relationship Outcomes,” *Journal of the Academy of Marketing Science*, 25(2), 127–37.

Rauschnabel, Philipp A. and Aaron C. Ahuvia (2014), “You’re so Lovable: Anthropomorphism and Brand Love,” *Journal of Brand Management*, 21, 372–95.

Reis, Harvey T. and Phillip Shaver (1988), “Intimacy as an Interpersonal Process.,” in *Handbook of Personal Relationships*, ed. S. W. Duck, John Wiley & Sons, 367–89, http://psycnet.apa.org/psycinfo/1988-97881-020.

Richardson, Daniel C and Rick Dale (2006), “Grounding Dialogue: Eye Movements Reveal the

Coordination of Attention During Conversation and the Effects of Common Ground," in *Proceedings of the Annual Meeting of the Cognitive Science Society, 28*, https://escholarship.org/uc/item/5m48273p.

Rijsdijk, Serge A., Erik Jan Hultink, and Adamantios Diamantopoulos (2007), "Product Intelligence: Its Conceptualization, Measurement and Impact on Consumer Satisfaction," *Journal of the Academy of Marketing Science*, 35(3), 340–56.

Schroll, Roland, Benedikt Schnurr, and Dhruv Grewal (2018), "Humanizing Products with Handwritten Typefaces," *Journal of Consumer Research*, 45, 648–72.

Sciutti, Alessandra, Martina Mara, Vincenzo Tagliasco, and Giulio Sandini (2018), "Humanizing Human-Robot Interaction," *IEEE Technology and Society Magazine*, (March), 22–29.

Searle, John H. (1969), *Speech Acts: An Essay in the Philosophy of Language*, Cambridge, MA: Cambridge University Press.

Serban, Iulian V., Chinnadhurai Sankar, Mathieu Germain, Saizheng Zhang, Zhouhan Lin, Sandeep Subramanian, Taesup Kim, Michael Pieper, Sarath Chandar, Nan Rosemary Ke, Sai Rajeshwar, Alexandre de Brebisson, Jose M. R. Sotelo, Dendi Suhubdy, Vincent Michalski, Alexandre Nguyen, Joelle Pineau, and Yoshua Bengio (2017), "A Deep Reinforcement Learning Chatbot," *ArXiv preprint*, 1–34.

Sprecher, Susan and Susan S. Hendrick (2004), "Self-Disclosure in Intimate Relationships: Associations with Individual and Relationship Characteristics over Time," *Journal of Social and Clinical Psychology*, 23(6), 857–77.

Sprecher, Susan, Stanislav Treger, Joshua D. Wondra, Nicole Hilaire, and Kevin Wallpe (2013), "Taking Turns: Reciprocal Self-Disclosure Promotes Liking in Initial Interactions," *Journal of Experimental Social Psychology*, 49(5), 860–66, http://dx.doi.org/10.1016/j.jesp.2013.03.017.

Sternberg, Robert J. and Susan Grajek (1984), "The Nature of Love.," *Journal of Personality and Social Psychology*, 47(2), 312–29.

Sundar, Shyam S., Xu Qian, and Saraswathi Bellur (2010), "Designing Interactivity in Media Interfaces: A Communications Perspective," in *CHI 2010: Perspectives on Design*, Atlanta, GA, USA, 2247–56.

Suri, Vipin K., Marianne Elia, and Jos van Hillegersberg (2017), "Software Bots -The Next Frontier for Shared Services and Functional Excellence," in *Global Sourcing of Digital Services: Micro and Macro Perspectives*, ed. I. Oshri, J. Kotlarsky, and L. Willcocks, Springer Cham, 81–94.

The Economist (2016), "Bots, the Next Frontier," *The Economist*, https://www.economist.com/business/2016/04/09/bots-the-next-frontier.

Thomson, Matthew (2006), "Human Brands: Investigating Antecedents to Consumers' Strong Attachments to Celebrities," *Journal of Marketing*, 70(3), 104–19.

Thomson, Matthew, Deborah J. MacInnis, and C. Whan Park (2005), "The Ties That Bind Measuring the Strength of Consumers," *Journal of Consumer Psychology*, 15(1), 77–91.

Thorbjørnsen, Helge, Magne Supphellen, Herbjørn Nysveen, and Per Egil Pedersen (2002), "Building Brand Relationship Online: A Comparison of Two Interactive Apllications," *Journal of Interactive Marketing*, 16(3), 17–34, http://dx.doi.org/10.1002/dir.10034.

Verhagen, Tibert, Jaap van Nes, Frans Feldberg, and Willemijn van Dolen (2014), "Virtual Customer Service Agents: Using Social Presence and Personalization to Shape Online Service Encounters," *Journal of Computer-Mediated Communication*, 19(3), 529–45.

Walker, Marilyn and Steve Whittaker (1995), "Mixed Initiative in Dialogue : An Investigation into Discourse Segmentation," *ArXiv preprint*.

Wang, Liz C, Julie Baker, Judy A Wagner, and Kirk Wakefield (2007), "Can a Retail Web Site Be Social?," *Journal of Marketing*, 71(3), 143–57, http://journals.ama.org/doi/abs/10.1509/jmkg.71.3.143.

Weger, Harry, Gina Castle Bell, Elizabeth M. Minei, and Melissa C. Robinson (2014), "The Relative Effectiveness of Active Listening in Initial Interactions," *International Journal of Listening*, 28(1), 13–31.

Wiemann, John M. and Mark L. Knapp (1975), "Turn-taking in Conversations," *Journal of Communication*, 25(2), 75–92.

Wright, Peter (2002), "Marketplace Metacognition and Social Intelligence," *Journal of Consumer Research*, 28(4), 677–82.

Wünderlich, Nancy V., Florian V. Wangenheim, and Mary J. Bitner (2013), "High Tech and High Touch: A Framework for Understanding User Attitudes and Behaviors Related to Smart Interactive Services," *Journal of Service Research*, 16(1), 3–20.

Xiaoyun Han, Robert J. Kwortnik, and Chunxiao Wang (2008), "Service Loyalty," *Journal of Service Research*, 11(1), 22–42.

Xu, Jingjun (David) (2016), "Retaining Customers by Utilizing Technology-Facilitated Chat: Mitigating Website Anxiety and Task Complexity," *Information and Management*, 53(5), 554–69, http://dx.doi.org/10.1016/j.im.2015.12.007.

Yadav, Manjit S. and Paul A. Pavlou (2014), "Marketing in Computer-Mediated Environments: Research Synthesis and New Directions," *Journal of Marketing*, 78(1), 20–40.

Yale, Laura and Alladi Venkatesh (1986), "Toward the Construct of Convenience in Consumer Research," in *Association for Consumer Research*, ed. Richard J. Lutz, Provo, UT, 403–8.

Zarouali, Brahim, Evert Van Den Broeck, Michel Walrave, and Karolien Poels (2018), "Predicting Consumer Responses to a Chatbot on Facebook," *Cyberpsychology, Behavior, and Social Networking*, 21(8), 491–97.

# 3. ARTICLE II

Bergner, A. (2020). Adaptive Sales Automation: Chatbots as Personalized and Scalable Sales Agents. *Marketing Review St. Gallen.*

**Abstract**

Artificial intelligence and natural language processing are fundamentally changing consumer–firm interactions. Sales automation via chatbots is often merely thought of as a cost-saving opportunity. This paper argues that well-designed chatbots allow to establish and nurture customer–firm relationships. The paper proposes a framework that assists in designing and dynamically adapting interface features to brand characteristics and customer needs. The framework is mapped along the traditional sales funnel, suggesting strategies for integrating chatbots throughout the entire customer journey.

*Keywords*: Chatbots, Sales Automation, Conversational Interfaces, Consumer-Brand Relationships

# Adaptive Sales Automation: Chatbots as Personalized and Scalable Sales Agents

## Management Summary

This paper proposes a framework for AI-powered sales automation with chatbots. The framework is mapped along the traditional sales funnel, suggesting strategies on how to better adapt these interfaces to the brand and specific customer needs.

## Main Propositions

- Chatbots provide an opportunity to augment a business's sales force through automation.
- Today, chatbots are primarily used to automate simple tasks rather than to establish and nurture personalized customer relationships.
- To leverage their full potential, chatbots should be designed to reflect the brand character and dynamically adapt to specific customer needs along the customer journey.

## Lessons Learned

- Design a chatbot with the same considerations in mind as you would when recruiting a sales agent to represent your firm.
- Use visual and verbal features of the interface to humanize the interaction and design a persona that accurately represents your brand.
- Collect data during an interaction to enable the chatbot to dynamically adapt to specific customer needs and buying tasks, as well as prepare for future interactions.
- Bring the chatbot outside its standard chat interface and allow it to interact with customers across various channels, integrating it into the full sales journey

In an age of sales automation (Shih 2016), what does it mean to automate personal selling? While some tasks such as lead scoring, visitor tracking, or campaign testing are considered quite simple to automate, the one-to-one trust relationship in customer service and sales is still largely regarded as a task only a human can fulfill (Buell 2018). Yet, advances in artificial intelligence and natural language processing are enabling a new kind of service encounter: chatbots. These humanized interfaces open up the possibility to emulate a one-to-one personalized sales encounter in a digital environment, while providing instant and "always on" customer support. Nevertheless, their success and adoption remain mixed (Hadi 2019). While industry reports indicate that 80% of businesses are planning to integrate chatbots by next year (BI Intelligence 2016), the vast majority of customers claim that they still prefer interacting with a human rather than with a chatbot (Press 2019). Thus, many businesses have so far considered chatbots merely a cost-saving opportunity to automate trivial online tasks rather than an opportunity to augment and upscale their sales force. As such, they are often designed as just another digital marketing tool rather than a personalizable touchpoint to nurture customer–brand relationships.

But what makes human interaction so unique? And how can chatbots be designed to better automate such an interaction? More specifically, what design features should businesses consider to turn these interfaces into valuable sales agents?

## Chatbots: The Status Quo

The implementation of chatbot interfaces varies tremendously in current industry practice, testimony to the very different roles these interfaces are fulfilling across businesses. As shown in Figure 1, some brands merely use a brand logo (e.g., Domino's), while others employ brand characters (e.g., Hipmunk's Raghav) or even human avatars (e.g., Amtrak's Julie) to interact with customers; some chatbots use formal filtering functions (e.g., Whole Foods Market), while others use visual product presentations (e.g., H&M) or a more conversational approach (e.g., Lidl's winebot Margot) to identify customer needs and personalize recommendations. More often than not, the implementation is a function of the default settings of the chatbot providers rather than a thoughtful process of how to design individual chatbot features to represent the brand and enable a meaningful customer interaction.

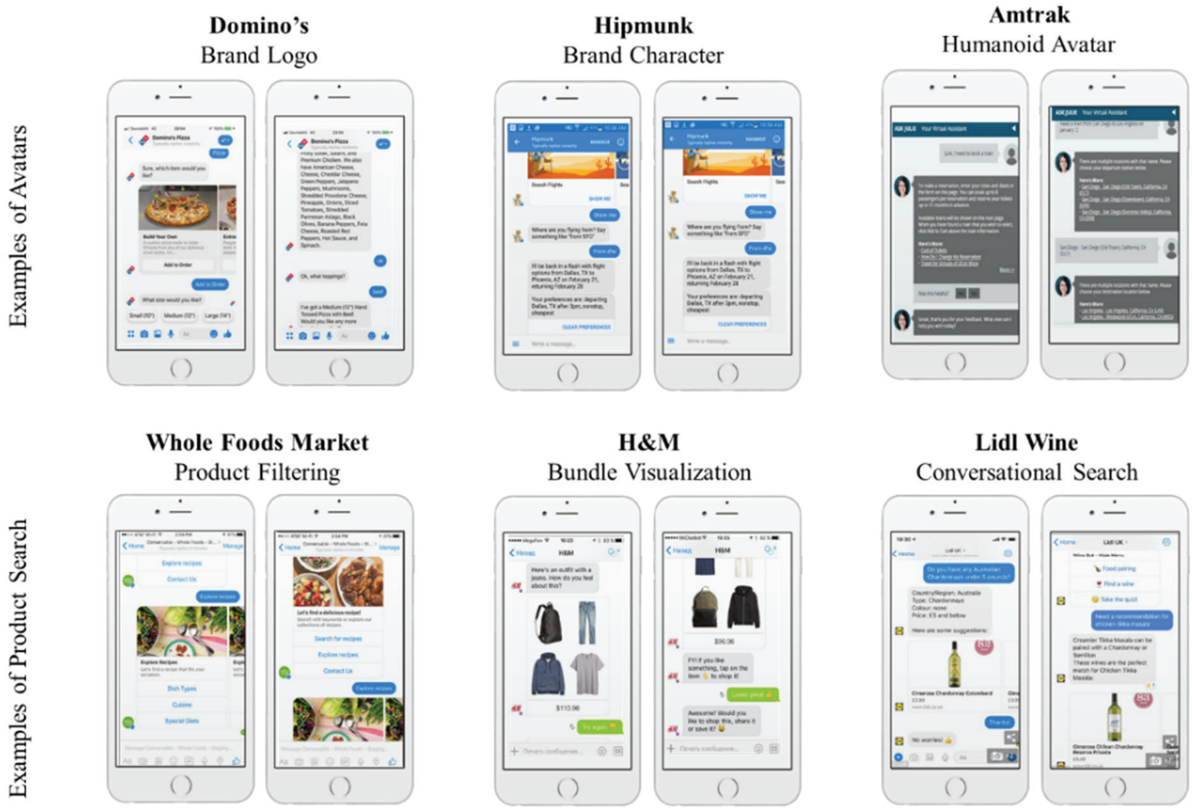



**Figure 3-1:** Examples of different chatbot implementations

**Adaptive Selling with Chatbots: An AI-Powered Sales Automation Framework**

Mapping chatbot interactions along the traditional sales funnel, there are four areas of adaptation for a chatbot interface:

1. First, an interface should be designed to reflect key brand characteristics and to be congruent with the brand image when targeting new customers, just like any other brand touchpoint (Haynes, Lackman, & Guskey, 1999; Park, Milberg, & Lawson, 1991). Given their unique features, chatbots provide brands with a highly scalable "human voice" that allows a one-to-one interaction with the customer. Thus, interface features should be carefully adapted to reflect brand traits and the brand's "personality" in order to create a consistent and relatable virtual spokesperson.

2. Second, during the interaction with a customer, chatbot interfaces should adapt to the user's unique characteristics and behaviors to create a more personalized experience tailored to the customer's needs. Building rapport through dialogue and a more personalized experience helps establish trust and nurture a more emotional connection with the brand (Bergner, Hildebrand, & Häubl, 2019; Hildebrand & Bergner, 2019).
3. Furthermore, chatbot interfaces should be adapted to contextual information in order to deliver the right interaction at the right time, based on where in the journey the customer currently is, as well as proposing to connect the customer with a human sales agent when needed.
4. Finally, a chatbot interface should not merely be considered an isolated touchpoint automating a single task, but rather be integrated into the full customer journey by proposing meaningful follow-ups designed to allow an ongoing interaction with the customer.

Figure 2 summarizes the critical steps that have to be mapped throughout the customer journey and the sales funnel. The specific chatbot design consideration at each step will be detailed in the following section.

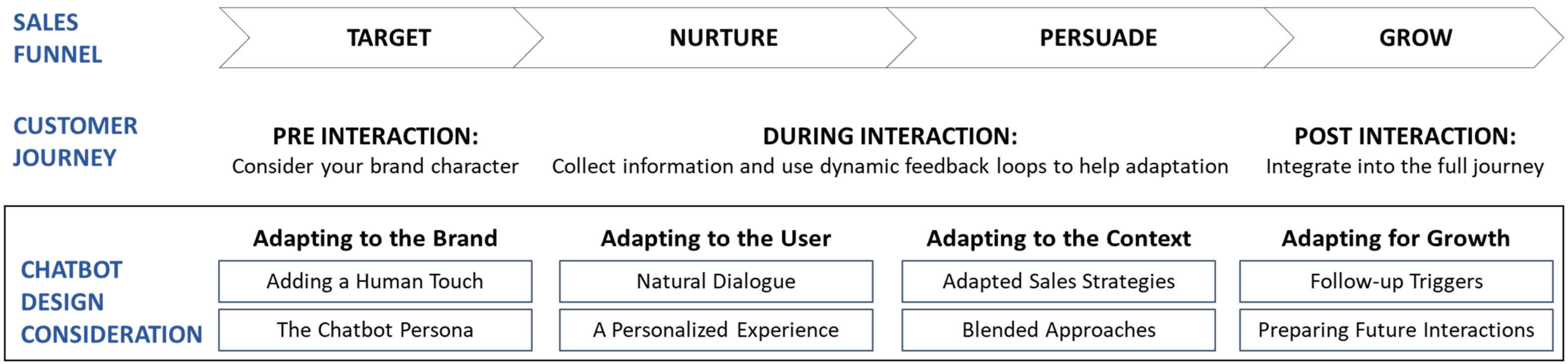



**Figure 3-2:** AI-powered sales automation framework

**Step 1: Adapting to the Brand**

*Adding a human touch*

Humanizing brands by imbuing them with unique traits and personalities is known to help build consumer–brand relationships (Aaker 1997; Fournier 1998) and drive sales (Holzwarth et al. 2006).

*Visual features.* In a chat interface, participants are generally represented by a profile image that can significantly influence perceived humanness and credibility of the interacting partner (Nowak and Rauh 2008). Aspects such as the gender or name impact how users interact with and relate to an interface (Araujo 2018; Waytz, Heafner, and Epley 2014). Brand characters or spokespersons, such as Hipmunk's chipmunk (see Figure 1), are known to humanize brands (Garretson and Niedrich 2004). Still, research has shown that people generally prefer interacting with more humanized rather than cartoon avatars (Nowak and Rauh 2006).

*Verbal features.* The ability of a chatbot interface to communicate in natural language makes the brand interaction "feel more human" compared to clicking on a static website, as language is considered a human capability (Jurafsky and Martin 2017). Language in humans is very rich in non-verbal cues and social conventions. In chatbot interfaces these unique aspects of human interaction can be implied by the use of emojis, which attribute human properties to the interface (Park & Sundar, 2015) and convey specific affective states (Dresner and Herring 2010). Furthermore, social conventions such as introductions, trivial acknowledgements and the ability to engage in small talk are all aspects unique to human interactions. The ability of an interface to engage in such behaviors therefore influences how human the interface is perceived to be (Bickmore and Cassell 2000). Figure 3 shows results from our own research, demonstrating that such visual and verbal features can significantly increase a chatbot's perceived humanness (Bergner, Hildebrand, and Häubl 2019).

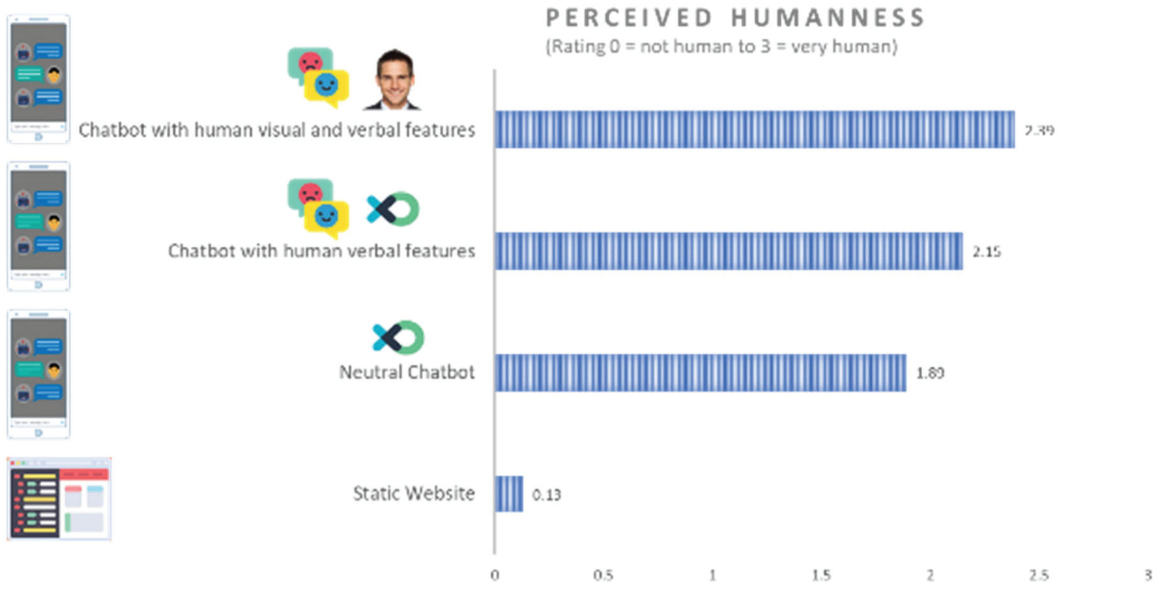



**Figure 3-3:** Perceived chatbot humanness

*The chatbot persona*

Aside from merely appearing human, chatbots should be considered brand advocates whose personality should be adapted to reflect key brand traits.

*Warmth vs. competence.* Just like humans, brand personalities are assessed along the two dominant dimensions of warmth and competence (Fiske et al. 2007). Warmth is related to friendliness, compassion, trustworthiness, and helpfulness, while competence is related to intelligence, skillfulness, expertise, and efficiency. An entity can be judged as both warm and competent, but warmth judgements generally happen faster and are predominantly based on affect (Fiske et al. 2007). How warm vs. competent a chatbot interface is designed to appear should be congruent with the brand personality. Visual cues such as facial expressions (Landwehr, Mcgill, and Herrmann 2011) and gender (Grohmann 2009), as well as more subtle cues in language and gesture can affect perceptions. For example, expressing empathy may increase a chatbot's perceived warmth, while demonstrating domain expertise may increase a chatbot's perceived competence.

*Extroversion.* A second aspect important to impression formation in interpersonal interactions is the extent of extroversion shown by the interacting parties. Extroversion has also been defined as the extent of "sociability" (Lucas et al., 2000) and is known to have prominent and distinct linguistic features (Hildebrand et al. 2020; Oberlander and Gill 2006). Specifically, a higher number of words, a certain level of abstraction and subjective language, or a preference for adjectives seem to be associated with more extrovert personalities (Beukeboom, Tanis, and Vermeulen 2013; Oberlander and Gill 2006). Thus, in a conversational interface, these syntactical features can provide important indications about the personality of the chatbot and the brand and should be carefully considered when designing chatbot responses.

**Step 2: Adapting to the User**

*Natural dialogue*

Before recent advances in artificial intelligence and natural language processing, dialogue has been an exclusive characteristic of human interaction. It is known to play an essential part in establishing trust and defining the social roles of interacting parties (Wiemann and Knapp 1975).

*Building rapport.* Research in linguistics and communication has revealed that turn-taking in interactions is a fundamental human characteristic, acquired even earlier than language itself (Levinson 2016). The act of taking turns during conversations has been shown to lead to more intimacy, liking and trust between interacting partners (Sprecher et al., 2013). A chatbot's ability to autonomously engage in such natural dialogic interaction is therefore key to improving engagement with the customer and establishing intimacy and trust. Its ability to react to various utterances and intents from the customer and engage in more trivial "social dialogue", such as answering unrelated questions, increases the number of turns and humanizes an interaction, thus influencing a customer's affect toward, and trust in, the chatbot and the brand (Bickmore and Cassell 2001a).

*Establishing social roles.* The rules and norms that govern a reciprocal turn-taking interaction (e.g., the amount of "air time" given to each party, or the knowledge of when to speak) are known to be key indicators with regard to the social roles of the interacting parties (Wiemann and Knapp 1975). Recent research on relationships with AI-enabled objects suggests that the social roles expressed by a smart object during an interaction may lead to fundamentally different customer experiences (Novak and Hoffman 2018). Thus, structural and verbal aspects of the dialogue should be adapted to reflect the chatbot's role during the interaction as an advisor, friend, or assistant. Using profiling questions, sentiment detection and feedback loops during the interaction can help to further adapt the chatbot's expressed social role to specific customer needs and expectations.

*A personalized experience*

Aside from the structural aspect of turn-taking, a second important aspect of interpersonal interaction is the extent of similarity and reciprocity between the interacting partners. By collecting information on the user before and during the interaction, the interface is able to adapt to user characteristics and mimic emotional expressions, creating a more intimate and personalized experience, similar to interacting with another human being.

*Similarity.* Similarity not only increases liking and creates a more intimate relationship, but can also have important downstream consequences on the effectiveness of persuasion attempts (Cialdini and Sagarin 1994). Adapting the visual features of the chatbot interface, such as the gender or name of the avatar, can create a more positive interaction for customers. Research on online avatars has shown that people tend to prefer interacting with avatars that match their own gender (Nowak and Rauh 2006). In our own research we found that personalizing chatbots to match customer characteristics significantly increased their persuasiveness when proposing trade-up options to customers (see Figure 4) (Bergner, Hildebrand, & Häubl, 2019).

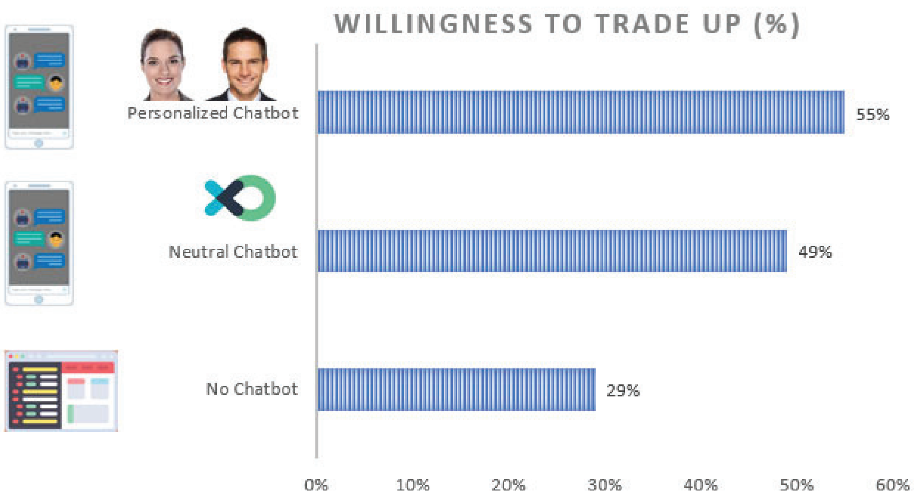



**Figure 3-4:** Consumers' willingness to trade-up depending on chatbot personalization

*Emotional reciprocity.* During interactions, humans tend to reciprocate emotions by mirroring non-verbal emotional expressions such as facial expressions, vocal features or body postures, which leads to more intimacy through a shared emotional experience called “emotional contagion” (Hatfield, Cacioppo, and Rapson 1993). While emotional reciprocity in text-based chatbot interfaces is limited, adapting subtle linguistic features and non-verbal cues to match customer behavior will create a more positive and intimate interaction. This requires that chatbots are not only able to mimic but also to interpret customers’ potentially ambiguous emotion expressions during conversations (e.g., use of emojis in place of words, use of abbreviations such as “lol”, etc.), in order to appropriately respond to them.

**Step 3: Adapting to the Context**

Research on effective sales strategies has shown that adapting to customer needs and the selling situation is key to driving sales (Franke and Park 2006; Spiro and Weitz 1990). Specifically, empirical findings show that such an individual customer orientation increases trust in the salesperson, nurtures brand relationships and increases overall customer satisfaction (Stock and Hoyer 2005).

*Adapted sales strategies*

*Buying task.* It is widely accepted that the buying task (e.g., new purchase vs. repurchases) has an important impact on the amount and type of information a customer requires. Specifically, for a new buy, customers may want to learn more about the brand, compare various products or ask a chatbot for guidance to determine the right product for their unique needs, while during a rebuy the chatbot may simply execute a previous order. By collecting and storing information on the purchase history of the customer, a chatbot can adapt the type and amount of information provided, as well as propose adapted up- or cross-selling opportunities to the customer.

*Buying stage.* The customer’s stage in the purchasing journey requires additional adaption of the interaction to help move the customer to the next stage in the purchase journey. Specifically, research on effective communication processes in sales has shown that successful salespeople adapt the time spent on task comments (e.g., providing information, comparisons, recommendations) and socio-emotional comments (e.g., support, agreement, acknowledgement) to the stage in the customer journey as well as individual customer traits

(Schuster and Danes 1986). By collecting information throughout the customer journey and creating ongoing feedback loops on the effectiveness of different communication strategies, the interface can dynamically adapt to such individual task needs.

*When to use a blended approach*

While this paper underlines the potential of chatbots as sales agents, there are certain situations that will require human intervention. Research has shown that anthropomorphism can backfire in the context of product wrongdoings or bad publicity (Puzakova, Kwak, and Rocereto 2013), as well as for dissatisfied customers (Hadi 2019). In their research, Hadi and colleagues show that a humanized customer service chatbot generally increased customer satisfaction, except for angry customers, for whom it drastically decreased satisfaction compared to interacting with a non-humanized chatbot (Hadi, 2019). One solution to this problem is a blended approach that redirects customers to a human sales agent when needed. By collecting information on the issue that needs to be resolved, as well as analyzing customer sentiment via natural language processing during the interaction (see, for example, Hildebrand et al., 2020), the chatbot can detect dissatisfied customers and swiftly connect them to a human counterpart better suited to solve the specific problem the customer is facing.

**Step 4: Adapting for Growth**

Finally, in order to augment the chatbot's impact as a sales agent for the brand, its interaction with the customer should not be considered a one-off isolated touchpoint, but rather be integrated with other communication channels of the brand.

*Follow-up triggers*

One way to keep the interaction going is by using automated follow-up triggers sent directly by the chatbot sales agent. For example, an automated follow-up email, personalized with customer information collected throughout the interaction, can help assess a customer's readiness to purchase or propose additional trade-up or trade-across opportunities. Another example would be a printed note accompanying a recent order, thanking the customer for the purchase in the name of the chatbot sales agent. Both procedures bring the chatbot outside of its standard medium of interaction and allow it to interact with the customer across the entire sales journey, just like a human sales agent would.

*Preparing for future interactions*

Data collection during the interaction with chatbots (e.g. via cookies, direct profiling questions, or using natural language processing; see Hildebrand and Schlager, 2019 for examples) is not only key to designing chatbots as truly adaptive user interfaces, but by storing this information, the interface can prepare future interactions with the customer more effectively. It can pick up a conversation where it was left off and provide the customer with the experience of a tailored one-to-one sales process.

## Discussion and Conclusion

This paper does not argue in favor of replacing human sales agents with chatbots. Rather, it proposes an augmentation of the sales force by bringing personal selling into the digital space through smart integration of chatbots into the customer journey. Chatbots need to be considered not merely a tool to automate isolated tasks, but rather an important brand advocate and a uniquely personalizable advisor. With recent developments in artificial intelligence, interacting with such humanized interfaces will potentially become indistinguishable from interacting with a human sales agent in the future. To leverage their full potential, brand managers need to carefully consider how to design these interfaces to reflect the unique personality of their brand while adapting to key customer needs throughout the sales journey. Only then can chatbots become highly valuable, scalable, and personalized sales agents for a firm, nurturing brand relationships and driving sales.

## References


Aaker, J. L. (1997). Dimensions of Brand Personality. *Journal of Marketing Research*, *34*(3), 347–356.

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.

Bergner, A., Hildebrand, C., & Häuble, G. (2019). Machine Talk: how conversational chatbots promote brand intimacy and influence consumer choice. *Society for Consumer Psychology Annual Conference*, 317–319. Savannah, USA.

Beukeboom, C. J., Tanis, M., & Vermeulen, I. E. (2013). The language of extraversion: extraverted people talk more abstractly, introverts are more concrete. *Journal of Language and Social Psychology*, *32*(2), 191–201. https://doi.org/10.1177/0261927X12460844.

BI Intelligence (2016). 80% of businesses want chatbots by 2020. Business Insider, December 14, 2016, https://www.businessinsider.com/80-of-businesses-want-chatbots-by-2020-2016-12?r=US&IR=T. Retrieved: November 13, 2018.

Bickmore, T., & Cassell, J. (2000). " How about this weather ?" Social Dialogue with Embodied Conversational Agents. *Proceedings of the American Association for Artificial Intelligence (AAAI) Fall Symposium on "Narrative Intelligence"*, Cape Cod, MA, 4–8.

Bickmore, T., & Cassell, J. (2001). Relational agents: a model and implementation of building user trust. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, *2*(2), 396–403. https://doi.org/10.1145/365024.365304.

Buell, R. W. (2018). The parts of customer service that should never be automated. Retrieved April 1, 2020, from Harvard Business Review, https://hbr.org/2018/02/the-parts-of-customer-service-that-should-never-be-automated.

Cialdini, R. B., & Sagarin, B. J. (1994). Principles of interpersonal influence. In T. C. Brock & M. C. Green (Eds.), *Persuasion: psychological insights and perspectives* (2nd ed., pp. 143–169). Thousand Oaks, CA: Sage Publications.

Dresner, E., & Herring, S. C. (2010). Functions of the nonverbal in CMC: emoticons and illocutionary force. *Communication Theory*, *20*(3), 249–268. https://doi.org/10.1111/j.1468-2885.2010.01362.x.

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

Fournier, S. (1998). Consumers and their brands: developing relationship theory in consumer research. *Journal of Consumer Research*, *24*(4), 343–353. https://doi.org/10.1086/209515.

Franke, G. R., & Park, J. E. (2006). Salesperson adaptive selling behavior and customer orientation: a meta-analysis. *Journal of Marketing Research*, *43*(4), 693–702. https://doi.org/10.1509/jmkr.43.4.693.

Garretson, J. A., & Niedrich, R. W. (2004). Spokes-characters: creating character trust and positive brand attitudes. *Journal of Advertising*, *33*(2), 25–36. https://doi.org/10.1080/00913367.2004.10639159.

Grohmann, B. (2009). Gender dimensions of brand personality. *Journal of Marketing Research*, *46*(1), 105–119. https://doi.org/10.1509/jmkr.46.1.105.

Hadi, R. (2019). When humanizing customer service chatbots might backfire. *NIM Marketing Intelligence Review*, *11*(2), 30–35.

Hatfield, E., Cacioppo, J. T., & Rapson, R. L. (1993). Emotional contagion. *Current Directions in Psychological Science*, *2*(3), 96–99.

Haynes, A., Lackman, C., & Guskey, A. (1999). Comprehensive brand presentation: ensuring consistent brand image. *Journal of Product & Brand Management*, *8*(4), 286–300. https://doi.org/10.1108/10610429910284229.

Hildebrand, C., Efthymiou, F., & Busquet, F. (2020). Voice analytics in business research: conceptual foundations, acoustic feature extraction, and applications. J*ournal of Business Research*, forthcoming.

Hildebrand, C., & Schlager, T. (2019). Focusing on others before you shop: exposure to Facebook promotes conventional product configurations. *Journal of Academy of Marketing Science*, *47*(2), 291–307.

Holzwarth, M., Janiszewski, C., & Neumann, M. M. (2006). The influence of avatars on online consumer shopping behavior. *Journal of Marketing*, *70*(4), 19–36. https://doi.org/10.1509/jmkg.70.4.19.

Jurafsky, D., & Martin, J. (2018). Dialog systems and chatbots. In: D. Jurafsky & J.H. Martin (Eds.), *Speech and Language Processing*. Draft online publication. https://web.stanford.edu/~jurafsky/slp3/26.pdf

Landwehr, J. R., Mcgill, A. L., & Herrmann, A. (2011). It' s got the look : the effect of friendly and aggressive "facial" expressions on product liking and sales. *Journal of Marketing*, *75*, 132–146. https://doi.org/10.2307/41228601.

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.

Lucas, R. E., Diener, E., Grob, A., Suh, E. M., & Shao, L. (2000). Cross-cultural evidence for the fundamental features of extraversion. *Journal of Personality and Social Psychology*, *79*(3), 452–468.

Novak, T., & 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*, 216–237. https://doi.org/10.2139/ssrn.3059093.

Nowak, K. L., & Rauh, C. (2006). The Influence of the avatar on online perceptions of anthropomorphism, androgyny, credibility, homophily, and attraction. *Journal of Computer-Mediated Communication*, *11*, 153–178. https://doi.org/10.1111/j.1468-2958.2006.00008.x.

Nowak, K. L., & Rauh, C. (2008). Choose your "buddy icon" carefully: the influence of avatar androgyny, anthropomorphism and credibility in online interactions. *Computers in Human Behavior*, *24*(4), 1473–1493. https://doi.org/10.1016/j.chb.2007.05.005.

Oberlander, J., & Gill, A. J. (2006). Language with character: a stratified corpus comparison of individual differences in e-mail communication. *Discourse Processes*, *42*(3), 239–270. https://doi.org/10.1207/s15326950dp4203_1.

Park, C. W., Milberg, S., & Lawson, R. (1991). Evaluation of Brand extensions: the role of product feature similarity and brand concept consistency. *Journal of Consumer Research*, *18*(2), 185. https://doi.org/10.1086/209251.

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.

Press, G. (2019). AI Stats News: 86% of consumers prefer humans to chatbots. Retrieved April 1, 2020, from Forbes Magazine, https://www.forbes.com/sites/gilpress/2019/10/02/ai-stats-news-86-of-consumers-prefer-to-interact-with-a-human-agent-rather-than-a-chatbot/#7e8c50c72d3b.

Puzakova, M., Kwak, H., & Rocereto, J. F. (2013). When humanizing brands goes wrong: the detrimental effect of brand anthropomorphization amid product wrongdoings. *Journal of Marketing*, *77*, 81–100. https://doi.org/10.1509/jm.11.0510.

Schuster, C. P., & Danes, J. E. (1986). Asking questions: some characteristics of successful sales encounters. *Journal of Personal Selling and Sales Management*, *6*(1), 17–27. https://doi.org/10.1080/08853134.1986.10754410.

Shih, C. (2016). Customer relationship automation is the new CRM. Retrieved April 1, 2020, from Harvard Business Review, https://hbr.org/2016/10/customer-relationship-automation-is-the-new-crm.

Spiro, R. L., & Weitz, B. A. (1990). Adaptive selling: conceptualization, measurement, and nomological validity. *Journal of Marketing Research*, *27*(1), 61. https://doi.org/10.2307/3172551.

Sprecher, S., Treger, S., Wondra, J. D., Hilaire, N., & Wallpe, K. (2013). Taking turns: reciprocal self-disclosure promotes liking in initial interactions. *Journal of Experimental Social Psychology*, *49*(5), 860–866. https://doi.org/10.1016/j.jesp.2013.03.017.

Stock, R. M., & Hoyer, W. D. (2005). An attitude-behavior model of salespeople's customer orientation. *Journal of the Academy of Marketing Science*, *33*(4), 536–552. https://doi.org/10.1177/0092070305276368.

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.

Wiemann, J. M., & Knapp, M. L. (1975). Turn-taking in conversations. *Journal of Communication*, *25*(2), 75–92. https://doi.org/10.1111/j.1460-2466.1975.tb00582.x.

# 4. ARTICLE III

Hildebrand, C., and 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.*

**Abstract**

The current work examines how automated financial advisory, in which consumers engage in a dynamic, dialogue-based interaction with a conversational robo advisor as opposed to static, non-conversational robo advisors, can alter perceptions of trust, the evaluation of a financial services firm, and consumer financial decision making. We develop and empirically test a novel conceptualization of conversational robo advisors building on prior work in human-to-human communication and interpersonal psychology, showing that conversational robo advisors cause greater levels of affective trust compared to non-conversational robo advisors and evoke a more benevolent evaluation of a financial services firm. We demonstrate that this increase in affective trust not only affects firm perception (in terms of benevolence attributions or a more positively-valenced onboarding experience), but has important implications for investor behavior, such as greater recommendation acceptance and an increase in asset allocation toward conversational robo advisors. These findings have important implications for research on trust formation between humans and machines, the effective design of conversational robo advisors, and public policy in the digital economy.

*Keywords*: Robo Advisors, Chatbots, Consumer Financial Decision Making, Investment Automation, Machine Intelligence, Trust.

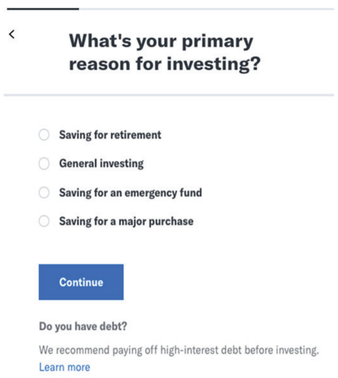
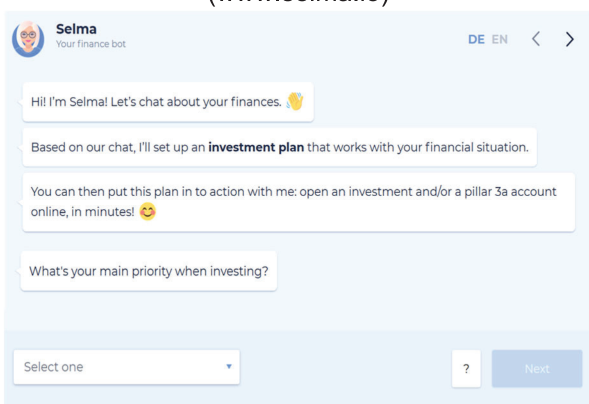
# Conversational Robo Advisors as Surrogates of Trust: Onboarding Experience, Firm Perception, and Consumer Financial Decision Making

Robo advisors have been praised as the next operating system in finance and the "new wealth management interface of the 21st century" (Andrus 2014). Robo advisors provide investment advice without the intervention of a human advisor. In short, robo advisors are digital interfaces that guide investors through an entirely automated process of investment advisory from assessing financial goals, evaluating consumers' risk profile, and ultimately managing the entire portfolio (Faloon and Scherer 2017; Gomber, Koch, and Siering 2017; Williams 2017). While discretionary input from consumers is possible, the key property is the fully automated process of risk assessment, asset allocation, and portfolio management, consistent with consumers' current financial situation, financial goals, and appetite for risk.

Despite the increasing presence of robo advisors in the financial industry with a share of $200 billion in assets under management worldwide in 2017 (Eule 2018) and companies such as Betterment or Wealthfront accumulating $1 billion in assets under management in less than 2.5 years after market entry (Moyer 2014), recent academic and industry studies reveal that growth rates are lacking behind expectations and that a broad consumer acceptance of robo advisors—even among young, affluent investors—has been surprisingly low (Jung et al. 2017; Schweitzer 2019). The majority of consumers still express a preference for human financial advisors, due to the lack of a "human touch" of robo advisors (Salmon 2018) and a human's greater ability to understand and personalize investment advice to consumers' unique financial situation (Hohenberger et al. 2019).

The current work introduces a novel conceptualization of robo advisors, which we refer to as "conversational robo advisors" using AI-enabled chatbots. We refer to conversational robo advisors as advisory interfaces that possess a dialogue-based process of financial advisory, which emulates fundamental properties of human-to-human conversations (such as turn-taking or the presence of social cues throughout a conversation; Davenport et al. 2020; Thomaz et al. 2020). The key hypothesis of the current research is that conversational robo advisors can provide an unexplored alternative to address and compensate for the lack of "human touch" during the advisory process compared to traditional, non-conversational robo advisors. Building on the effective properties of human-to-human conversations in prior communication research and interpersonal psychology (Fiske, Cuddy, and Glick 2007; Levinson 2016; Sprecher et al. 2013b), we develop and empirically test our conceptualization of "conversational robo advisors" and how they affect consumer perceptions of trust, firm evaluation, and investor behavior. Table 1 provides a comparison of the key conceptual features contrasting non-conversational robo advisors (i.e., possessing static, self-report, and one-way communication features) compared to conversational robo advisors (i.e., possessing dynamic, dialogue-based, and turn-taking communication features).

**Table 4-1:** Key conceptual features of non-conversational vs. conversational robo advisors

| Non-Conversational Robo Advisor | Conversational Robo Advisor |
|---|---|
| *Example* | |
| Betterment (www.betterment.com) | Selma Finance (www.selma.io) |
|  |  |
| *Key Features* | |
| Static | Dynamic |
| Self-Report | Dialogue-Based |
| One-Way | Turn-Taking |
| Formal Language (few social cues) | Informal Language (rich social cues) |
| Absent Anthropomorphic design cues | Present Anthropomorphic design cues |





Across four studies, we provide empirical evidence that conversational as opposed to non-conversational robo advisors evoke greater levels of affective trust toward a robo advisor, and that these greater levels of trust in turn alter firm perception and investor behavior. Specifically, we demonstrate that conversational as opposed to non-conversational robo advisors increase consumers' likelihood to follow portfolio recommendations, positively affect consumers' attributions of benevolence toward a financial services firm (believing that a financial services firm acts in one's best interest), and provide more engaging, positively-valenced advisory experiences for consumers.

In what follows, we first review prior work on robo advisory and the relationship between automation and trust. We then develop a set of key hypotheses on how conversational robo advisors alter attributions of trust, firm perception, and investor behavior, and present the results of four studies designed to test our theorizing. We conclude with a discussion of the current findings for research on automation, trust, and public policy in the digital economy.

## Theory and hypotheses

*Theoretical background and literature review*

Robo advisors are digital interfaces that guide private investors through an entirely automated process of investment advisory (Gomber et al. 2017). The predominant mode of robo advisory employs a static, formalized process of self-reports to assess consumers' financial situation and individual risk profile (see Table 1 for example). These self-reports are generally used to gather information on investors' goals, existing financial assets, and appetite for risk (Tedesco 2015), which are then translated into an adequate portfolio of financial investments that can be managed automatically by the advisory system (Faloon and Scherer 2017).

As summarized in the literature review of Table 2, previous research on financial robo advisors has mainly focused on designing algorithms with respect to the "optimal" composition of asset classes based on consumers' risk profile and financial goals (Day, Lin, and Chen 2018; Kilic, Heinrich, and Schwabe 2015; Musto et al. 2015), usability aspects of the interface (Jung et al. 2018), or the interactive role of the company's sales channel and profit orientation to rely on robotic investment advice (Lourenço, Dellaert, and Donkers 2020). Robo advisors have been considered predominantly as static tools to replace and optimize the human advisory process (see Lourenço, Dellaert, and Donkers 2020 for a notable exception, using an interactive "pension builder" interface), to collect client data and allocate an appropriate portfolio consistent with a clients' risk profile, rather than leveraging the relationship-building potential for the financial services firm or the opportunity to more directly address the lack of broader acceptance of financial robo advisory and lack of a "human touch."

At the same time, recent advances in artificial intelligence and natural language processing have given rise to conversational interfaces, or so called AI-enabled "chatbots" (Dale 2016; Davenport et al. 2020; Thomaz et al. 2020), that use a retrieval-based dialogue system that emulates the characteristics of human-to-human conversations. Thus, a second generation of robo advisors is emerging which we refer to as "conversational robo advisors." The dialogue-based interaction modality of conversational robo advisors possesses key features (see Table 1), whose malleability enable a more personalized, one-to-one interaction between the investor and financial services firm. First, the interaction is based on a dynamically construed dialogue flow that emulates the characteristics of a human-to-human conversation through a sequential process of turn-taking. Thus, rather than a static, questionnaire-like request for information as in non-conversational robo advisors, the interaction with a conversational robo advisor is based on the structural aspects of a two-sided conversation. Second, the language used by the interface can be semantically adapted to display richer social cues, such as emojis or trivial acknowledgments to signal emotions and active listening (Thomaz et al. 2020). Finally, the conversational interface can further be humanized by integrating anthropomorphic design cues into the interface and altering the visual appearance through avatars or other forms of visual display (Araujo 2018).

Recent research has started to investigate the impact of some of these malleable factors in conversational interfaces, showing that anthropomorphic design cues can positively impact service satisfaction and firm performance (Adam et al. 2019; Köhler et al. 2011), while research on so called "conversational sales agents" in the online retail domain revealed that interfaces with human-like cues tend to foster greater trust in and likability of a sales agent (Araujo 2018; Luo et al. 2006), mimicking related research highlighting the importance of creating a stronger social presence in human–robot relationships (van Doorn et al. 2017; Mende et al. 2019). Building on and integrating these distinct literature streams on financial advisory and conversational agents, the key objective of the current research is to provide a causal test of whether "conversational robo advisors" may alter consumers' experience of the advisory process, firm perception, and ultimately investor behavior. In what follows, we develop a set of key hypotheses along with an overview of the empirical studies designed to test these hypotheses.

**Table 4-2:** Review of relevant literature

| Study | Interface Type | Context | Task & Stimuli | Theoretical Background | Independent Variable | Psychological Process | Dependent Variable | Main Findings |
|---|---|---|---|---|---|---|---|---|
| Jung, Dorner, Weinhardt, & Pusmaz (2018) | Non-conversational | Robo Advisor | Interactive Portfolio Selection with Robo Advisor | Design Science | Ease of Interaction | Transparency | Trust | Ease of interaction signals transparency which leads to higher customer trust. |
| Musto, Semeraro, Lops, De Gemmis, & Lekkas (2015) | Non-conversational | Robo Advisor | Asset Allocation with OFS Advice Wealth-Management Platform | Case-Based Reasoning | Framework Parameters | . / . | Performance of Robo Advisor (vs. Human Advisor) | Robo advisors outperform human advisors while meeting the preferred risk profile. |
| Touré-Tillery, & McGill (2015) | Non-conversational | Shopping Assistant | Watch Online Advertisements for Dental Floss, Lightbulb, Coffee Mug | Anthropo-morphism | Anthropo-morphism Interpersonal Trust | Attention Goodwill | Persuasiveness | People low in interpersonal trust are more persuaded by anthropomorphized messengers than by human spokespeople. |
| Luo, McGoldrick, Beatty, & Keeling (2006) | Non-conversational | Shopping Assistant | Watch Scenario of Online Booking Process | Character Design | Gender Facial Appearance (human vs. cartoon) | . / . | Likability Trust | Human-like characters are perceived as more trustworthy than cartoon like characters. |
| Adam, Toutaoui, Pfeuffer, & Hinz, (2019) | Conversational | Robo Advisor | Interactive Onboarding and Portfolio Selection with Robo Advisor | Anthropo-morphism Anchoring | Anthropo-morphism | Social Presence | Investment Volume | Anthropomorphized robo advisors lead to higher perceptions of social presence and higher investment volume. |
| Köhler, Rohm, de Ruyter, & Wetzels (2011) | Conversational | Online Banking Assistant | Survey of Self-Selected Customers Using Online Banking Assistant | Social Learning Theory | Interaction Style (proactive, reactive) Content (social, functional) | . / . | Account Performance Newcomer Adjustment | Proactive (vs. reactive) interaction style paired with delivery of functional (vs. social) content improves newcomer adjustment (service use efficacy of new customers), which in turn improves firm level performance. |
| Araujo (2018) | Conversational | Shopping Assistant | Interactive Flower Ordering Task with Chatbot | Embodiment | Anthropo-morphism Intelligence-Framing | Social Presence | Emotional Connection Service Satisfaction | Conversational interfaces that use more human-like cues build a stronger emotional connection with the company, independent of whether a chatbot is framed as intelligent or not. |
| This research | Non-conversational vs. Conversational | Robo Advisor | Interactive Risk Assessment & Portfolio Selection with Non-Conversational vs. Conversational Robo Advisor | Speech Formation Turn-Taking | Turn-Taking Social Cues | Affective Trust | Firm Perception (Benevolence Attribution) Investor Behavior (Recommendation Acceptance, Portfolio Allocation) | Turn-taking and social cues lead to higher affective trust in a robo advisor, more positive firm perception, and greater recommendation acceptance (even if objectively wrong or invoking larger annual fees). |

*Turn-taking and affective trust*

The critical feature of conversational robo advisors is their capacity to take turns during the initial onboarding phase. Such turn-taking mimics natural iterations in human-to-human conversations which have been shown to act as an inherent trust-building mechanism (Bickmore and Cassell 2001b). Although healthy, pre-existing relationships exert this property naturally, experimental evidence suggests that turn-taking can effectively contribute to the formation of a more affective, trusting relationship (Sprecher et al. 2013b). Specifically, turn-taking during conversations among humans and a more equal share of "air time" is linked to better relational outcomes and greater perceptions of trustworthiness (Levinson 2016; Wiemann and Knapp 1975). For example, inviting mutually unfamiliar individuals to a 12 minute conversation using a pre-defined turn-taking protocol has been shown to maximize liking, closeness, and enjoyment of a conversation partner compared to non-reciprocal dyads that were not requested to take turns during a conversation (Sprecher et al. 2013b).

The origin of this mechanism lies deep in human development, stemming from a combination of instinctive and learned behavior (Levinson 2016), representing social and cooperative efforts between humans (Grice 1975). The process of turn-taking is often further governed by both verbal and non-verbal social cues, indicating active listening such as providing trivial acknowledgements of what the conversation partner just said or implicit signals to indicate whether the speaker is ready to yield the turn or whether an answer is expected from the listener (Wiemann and Knapp 1975). This back-and-forth communication protocol is an essential trust-building mechanism in human-to-human interactions. Indeed, even trivial acknowledgments or interludes of "small talk" can signal greater involvement and understanding from the side of the interaction partner and build stronger rapport (Bickmore and Cassell 2000; Cappela 1985). Thus, turn-taking is not only a process defining a dialogue-based interaction (i.e., how the conversation is structured), but is also an inherently social activity with both verbal and non-verbal social cues (i.e., acknowledgements and verbal affirmations such as "Got it!" or "I see" as well as non-verbal affirmations such as emotional displays or body movements such as head nodding to indicate active listening), all of which are shaping the social roles and relationships of the interacting parties. Furthermore, prior work on embodied interfaces has shown that users generally feel more comfortable and prefer to interact with systems that are capable of turn-taking, and that adherence to this unique conversational protocol is at least as important as other factors, such as the actual display of emotions by the interface (Cassell and Thórisson 1999). This is consistent with prior work in human-robot interactions, showing that more social behaviors, such as turn-taking, contribute to the formation of trust and are associated with a greater willingness to interact with the same robot in the future (Looije, Neerincx, and Cnossen 2010).

Thus, we expect that the inherent turn-taking capacity of conversational as opposed to non-conversational robo advisors enhances affective levels of trust toward the robo advisor. Affective trust is a distinct measure of relational trust between two parties, differing from other forms of trust (such as cognitive trust, which focuses on the objective assessment of competence and quality dimensions of an interaction partner; Johnson and Grayson 2005). Affective trust is a more emotional, subjective dimension of trust, linked to the social nature of a relationship. Taken together, we expect that the turn-taking capacity of conversational robo advisors increases perceptions of affective trust relative to non-conversational robo advisors.

**H1:** Conversational as compared to non-conversational interfaces lead to greater levels of affective trust in the robo advisor.

*Firm perception and investor behavior*

Greater levels of affective trust have been shown to positively affect a wide array of both perceptual and behavioral outcomes, from increasing relationship satisfaction and team performance (Jones and George 1998) to cooperative behavior (Rousseau et al. 1998) and greater long-term customer loyalty (He, Li, and Harris 2012). Most importantly, affective trust is a critical ingredient when facing decisions that involve risk (Pavlou 2003). Building on our theorizing that conversational versus non-conversational interface might cause greater affective trust, we expect subsequent changes in firm perception and investor behavior in response to these enhanced levels of affective trust. First, developing greater affective trust in one situation has been shown to generate the belief that the other entity is willing to act in the best interest of the focal individual in a subsequent task (Cho 2006; Johnson and Grayson 2005). Thus, we hypothesize that greater perceptions of affective trust in turn stimulates subsequent attributions of benevolence, i.e. the assumption that another entity might act in one's best interest (Xie and Peng 2009). This is particularly important in the context of financial institutions in which consumers often expect the financial institution to act in its own self-interest (Monti et al. 2014). Thus, we hypothesize that greater levels of affective trust toward the robo advisor may generate subsequent attributions of benevolence toward the financial services firm to act in a client's best interest.

**H2:** Greater levels of affective trust in a conversational as comparted to a non-conversational robo advisor lead to higher attributions of benevolence toward a financial services firm.

Greater affective trust may also change investor behavior more directly. Greater affective trust has been shown to make people more susceptible to influence attempts (Aronson 1999; Fransen et al. 2015) and to elevate persuasion effectiveness more broadly (Betancourt 1990; Hexmoor et al. 2008). For example, greater levels of trust in a bank have been found to elevate clients' use of more uncertain transaction infrastructures (Yousafzai et al. 2010). Thus, we expect that increasing consumers' levels of affective trust in turn increases their willingness to follow or accept the financial advice received, such as choosing a recommended financial portfolio.

**H3:** Greater levels of affective trust in a conversational as compared to a non-conversational robo advisor increases consumers' willingness to accept a recommended financial portfolio.

In summary, the current research tests three key predictions of how conversational as opposed to non-conversational robo advisors change consumer financial decision making and the downstream consequences of these decisions. First, we hypothesize that conversational robo advisors evoke greater levels of affective trust. Second, we predict that increased levels of affective trust caused by conversational as opposed to non-conversational robo advisors lead to a more benevolent perception of the financial services firm. Finally, we expect that the increased levels of affective trust caused by conversational as opposed to non-conversational robo advisors lead to higher recommendation acceptance.

**Overview of studies and experimental paradigm**

Study 1 tests our baseline hypothesis of whether conversational robo advisors cause greater levels of affective trust compared to non-conversational robo advisors, and whether enhanced levels of affective trust spill over to a more benevolent evaluation of a financial services firm. Study 2 provides evidence that greater affective trust not only enhances firm perception (attribution of benevolence toward the financial services firm) but also influences behavioral outcomes (asset allocation toward conversational as opposed to non-conversational robo advisors). Studies 3 and 4 further extend these findings by showing that greater affective trust increases consumers' likelihood to accept portfolio recommendations even if they are objectively wrong (Study 3) or if they invoke larger annual fees (Study 4).

**Study 1**

Study 1 was designed to provide a first test of our theory on whether conversational as opposed to non-conversational robo advisors alter perceptions of affective trust and whether these changes in affective trust trigger the predicted increase in benevolence attributions toward a financial services firm.

*Design and procedure*

A total of 307 participants were recruited from a nationwide online consumer panel (Prolific) and preselected based on their interest in financial services and being active private investors ($M_{Age}$ = 43.79, $SD_{Age}$ = 15.20, 52% females). Participants were randomly assigned to one of three conditions: (1) a non-conversational interface condition, (2) a conversational interface condition without social cues, or (3) a conversational interface condition with social cues (see Figure 1 for an illustration of the three robo advisor conditions). Across all conditions, participants went through a risk profiling assessment involving their current financial situation, their financial goals, and their perception of financial risk (see Web Appendix A.1). The sequence and format of all questions was identical across conditions. The two conversational robo advisor conditions differed in the presence versus absence of social cues and were otherwise identical in the extent of turn-taking or sequence of risk profiling questions. In the conversational interface condition with social cues, the robo advisor provided trivial acknowledgements (e.g., "Great, thanks!") and displayed emoticons during the conversation (see Figure 1). In the conversational interface condition without social cues, the interface was identical relative to the non-conversational interface condition except for the key conceptual difference in response and expression modality (i.e., using a chat console during the advisory process but holding both the sequence, content, and all other features of the interaction constant).

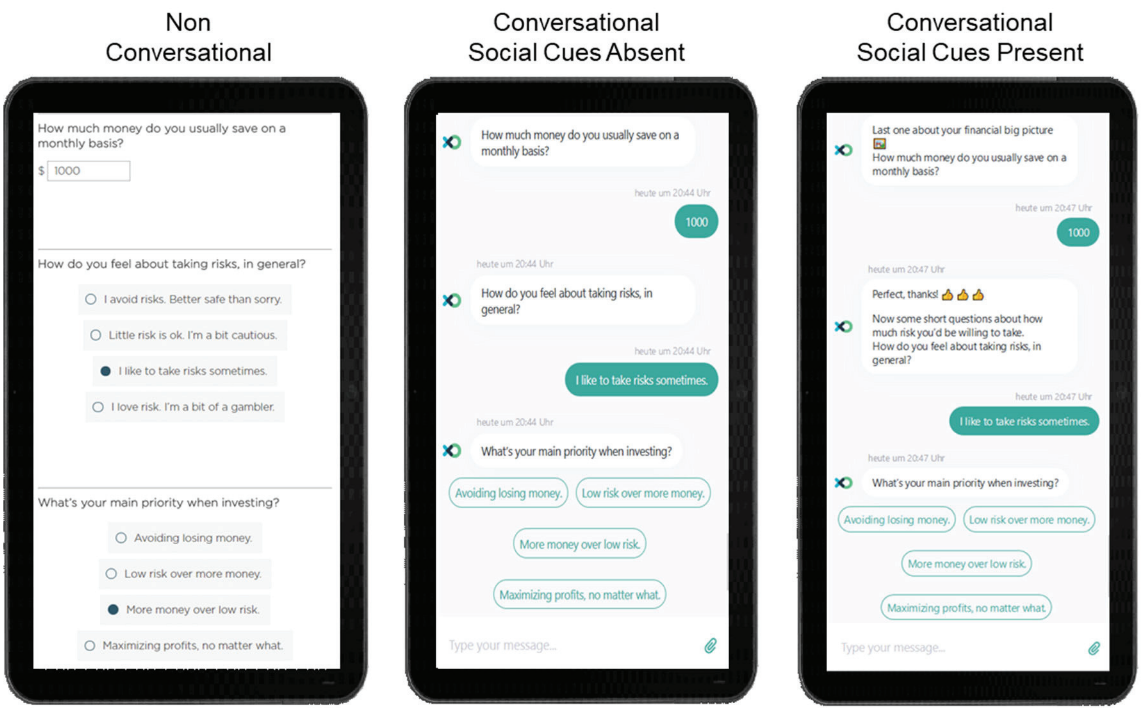



**Figure 4-1:** Exemplary robo advisor interface conditions

*Measurement*

Immediately after participants completed the robo advisory task, they were forwarded to the original questionnaire and we assessed their level of affective trust toward the robo advisor using three items (scale adapted from Johnson and Grayson 2005, sample item: "This financial advisory system displayed a warm and caring attitude toward me"; 7-point Likert scale, from 1:"I do not agree at all" to 7:"I fully agree"; $\alpha_{TrustAff} = .90$, see Web Appendix A3.1 and A5 for further details on scale consistency). Next, we measured participants' attribution of benevolence toward the financial services firm using 5 items (scale adapted from Schlosser et al. 2006: "This advisory company seems very concerned about my welfare"; 7-point Likert scale, from 1: "I do not agree at all" to 7:"I fully agree"; $\alpha_{FirmBenevolence} = .92$, see Web Appendix A3.2 and A5).

To generate additional qualitative insight, we used an open response technique at the end of the study to further assess consumers' spontaneous thoughts about the interaction with the financial advisory system. Participants first wrote down their spontaneous thoughts and feelings related to the robo advisory interface they used and then evaluated the valence of each of these thoughts as negative (coded as -1), neutral (coded as 0), or positive (coded as 1). All open text responses of the thought-listing task were concatenated into a single text vector. We used the AFINN lexicon and the tidytext package in R (Nielsen 2011) for all subsequent text processing and sentiment analyses.

*Pretest*

We conducted a pre-test with a total of 178 participants ($M_{Age} = 34.41$, $SD_{Age} = 9.86$, 30% females) from the same population and using the same selection criteria as in the main study to assess the effectiveness of the experimental manipulation. Specifically, the objective of this pre-test was to assess whether the two conversational robo advisor conditions compared to the non-conversational robo advisor condition (1) were perceived as greater in the extent of turn-taking and (2) whether the conversational interface condition with social cues was perceived as richer in the extent of social cues compared to conversational interface condition without social cues and compared to the non-conversational interface condition. Participants were randomly assigned to the same three experimental conditions as in the main study (non-conversational robo advisor interface vs. conversational robo advisor interface with vs. without social cues) and completed the same risk profiling questionnaire as in the main study. Immediately after completing the risk profiling questionnaire, we assessed participants' perceived degree of turn-taking (adapted from Song and Zinkhan 2008; sample item: "The system facilitated a two-way conversation," $\alpha_{TurnTaking}= .91$, see Web Appendix A3.3) and the perceived presence of social cues (sample item: "The financial advisory system … displayed smiling faces and other emoticons during the task / acknowledged and confirmed my input (e.g., 'Great, thanks!')," $\alpha_{SocialCues} = .85$, see Web Appendix A3.4). Supporting the effectiveness of the experimental manipulation, the interface manipulation had a significant effect on turn-taking ($F(2, 175) = 12.61$, $p < .001$) and perception of social cues ($F(2, 175) = 15.79$, $p < .001$). Follow-up contrasts confirmed that participants in both conversational interface conditions perceived the extent of turn-taking as significantly higher compared to the non-conversational interface condition ($M_{Conversational\_SocialCuesAbsent} = 5.21$, $M_{NonConversational} = 4.28$, $t = 3.975$, $p < .01$; $M_{Conversational\_SocialCuesPresent} = 5.37$, $t = 4.691$, $p < .001$), while we found no significant difference between both conversational interface conditions ($t = 0.732$, $p > .46$). Furthermore, participants in the social cues present condition perceived a significantly larger number of social cues compared to both the non-conversational interface condition ($M_{Conversational\_SocialCuesPresent} = 5.31$, $M_{NonConversational} = 3.59$, $t = 5.566$, $p < .001$) and the conversational interface condition without social cues ($M_{Conversational\_SocialCuesAbsent} = 4.28$, $t = 3.404$, $p < .01$), while the difference between the non-conversational interface condition and the conversational interface without social cues was marginally significant ( $t = 2.235$, $p = .07$). Taken together, these findings provide support for the effectiveness of the experimental manipulation indicating that both conversational interface conditions were perceived as greater in the extent of turn-taking compared to a non-conversational interface and that the conversational interface with social cues possessed a significantly higher presence of social cues during the advisory task.

*Results*

*Affective trust toward advisor.* The interface manipulation had a significant effect on perceptions of affective trust ($F(2, 304) = 11.740$, $p < .001$). Specifically, follow-up contrasts with Holm correction for family-wise errors revealed that participants attributed a significantly greater level of affective trust toward the conversational interface without social cues compared to the non-conversational interface ($M_{Conversational\_SocialCuesAbsent} = 3.90$, $M_{NonConversational} = 3.27$, $t = 2.750$, $p < .05$).

This effect further increased in the conversational interface with social cues condition, both when comparing to the non-conversational interface ($M_{Conversational\_SocialCuesPresent}$ = 4.37, $t$ = 4.834, $p$ < .001), as well as to the conversational interface without social cues ($t$ = 2.093, $p$ < .05).

*Benevolence attribution toward firm*. Similarly, we found a significant effect of the type of interface on attributions of benevolence toward the financial services firm ($F(2, 304) = 8.259$, $p < .001$). Mirroring the results of affective trust, follow-up contrasts with Holm correction confirmed that participants who interacted with the conversational interface without social cues perceived the financial services firm as significantly more benevolent than participants who interacted with the non-conversational interface ($M_{Conversational\_SocialCuesAbsent}$ = 3.75, $M_{NonConversational}$ = 3.22, $t$ = 2.627, $p$ < .05). This effect further increased in the conversational interface with social cues compared to the non-conversational interface ($M_{Conversational\_SocialCuesPresent}$ = 4.03, $t$ = 4.006, $p$ < .001). The difference between both conversational interface conditions was non-significant ($t$ = 1.384, $p$ = .17).

*Mediation.* Next, we estimated a simple mediation model to provide a direct test of our theorizing. Using effect contrasts of the interface condition (coding the non-conversational robo advisor as -1, the conversational robo advisor without social cues as 0, and the conversational robo advisor with social cues as 1), we estimated a simple mediation model (5000 bootstrap samples) in which the interface condition served as the independent, affective trust as the mediator, and perceived benevolence of the financial services firm as the dependent variable. In support of our theorizing, the positive main effect of the conversational interface on attributions of benevolence toward the financial services firm ($\beta_{Conversational}$ = .37, $t$ = 3.124, $p$ < .01) was significantly mediated via first enhancing affective trust in the robo advisor ($\beta_{TrustAff}$ = .64, $t$ = 17.926, $p$ < .001), rendering the residual direct effect non-significant ($\beta_{Conversational\_Direct}$ = .03, $t$ = .342, $p$ = .73), indicating full mediation ($\beta_{Indirect}$ = .34, 95% CI: [.17; .51]).

*Sentiment analysis*. To gain additional qualitative insight into consumers' thoughts and feelings, we analyzed the thought listing task as follows. As highlighted in the methods and procedure section, we concatenated participants' thoughts into a single text vector and assessed sentiment valence based on participants' polarity ratings. Providing additional evidence for our theorizing, these results reveal a significant effect of interface type on sentiment valence ($F(2, 304) = 13.500$, $p < .001$). Follow-up contrasts with Holm correction for family-wise errors confirm that participants who interacted with the conversational interface without social cues evaluated the robo advisor interface more positively than those who interacted with the non-conversational interface ($M_{Conversational\_SocialCuesAbsent} = .07$, $M_{NonConversational} = -.19$; $t = 2.959$, $p < .01$). This effect further increased in the conversational interface with social cues compared to both the non-conversational interface ($M_{Conversational\_SocialCuesPresent} = .26$, $t = 5.183$, $p < .001$) and the conversational interface without social cues ($t = 2.234$, $p < .05$).

To illustrate the pattern of results based on these qualitative insights, we analyzed participants' qualitative responses using the AFINN lexicon in the tidytext package in R. Figure 2 demonstrates the positive relationship between affective trust and firm benevolence (i.e., positive slope across all conditions). Each observation represents the most frequent sentiment per participant and is further color-coded based on the corresponding sentiment valence (using the AFINN lexicon). These findings illustrate that greater affective trust is associated with a more positive evaluation of the financial services firm and a greater density of positively valenced associations in the conversational interface conditions (such as feeling excited, joyful, and interested) as opposed to the non-conversational interface condition with more negatively valenced sentiments (such as feeling bored or stressed).

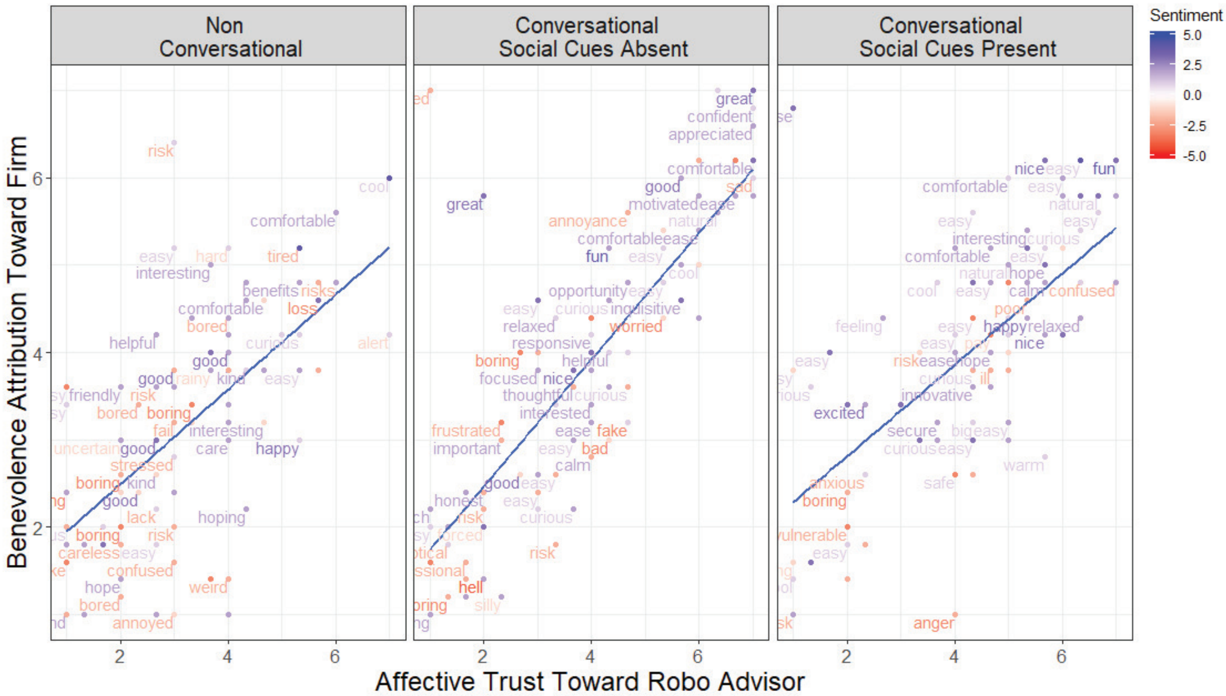



**Figure 4-2:** Affective trust, benevolence, and consumer sentiment

*Discussion*

The findings of Study 1 provide evidence that both conversational robo advisors (with or without social cues) evoke greater levels of affective trust as opposed to non-conversational interfaces. These changes in the interface modality in turn led to a more positive firm evaluation, and a more positively valenced advisory experience for consumers. Providing additional social cues led to overall larger effect sizes when contrasting both conversational interface conditions. Based on these initial findings and the increasingly dominant trend to anthropomorphize conversational interfaces across industries (BenMark and Venkatachari 2016), all subsequent studies will build on the conversational interface condition with present social cues during the advisory process.

## Study 2

Can conversational as opposed to non-conversational interfaces also affect investor behavior more directly? The key objective of Study 2 was to test whether private investors are willing to let a conversational as opposed to non-conversational robo advisor manage their assets as part of an incentive-compatible investment game.

*Design and procedure*

A total of 309 participants participated in this study using the same preselection criteria as in Study 1 ($M_{Age}$ = 36.32, $SD_{Age}$ = 10.48, 36% females). Participants were randomly assigned to either a non-conversational interface condition or a conversational interface condition with social cues. As in Study 1, all conditions involved the same risk profiling task in exactly the same sequence. At the end of the risk profiling task, participants received a personalized investment portfolio based on their risk profile, displaying the percentage of asset classes such as bonds, high yield bonds, stocks, and cash, with either a capital preserving (only 5% stocks), balanced (30% stocks), or growth-oriented investment portfolio (85% stocks) (see Web Appendix A1.2 and A2.1 for a detailed description of the portfolio matching algorithm).

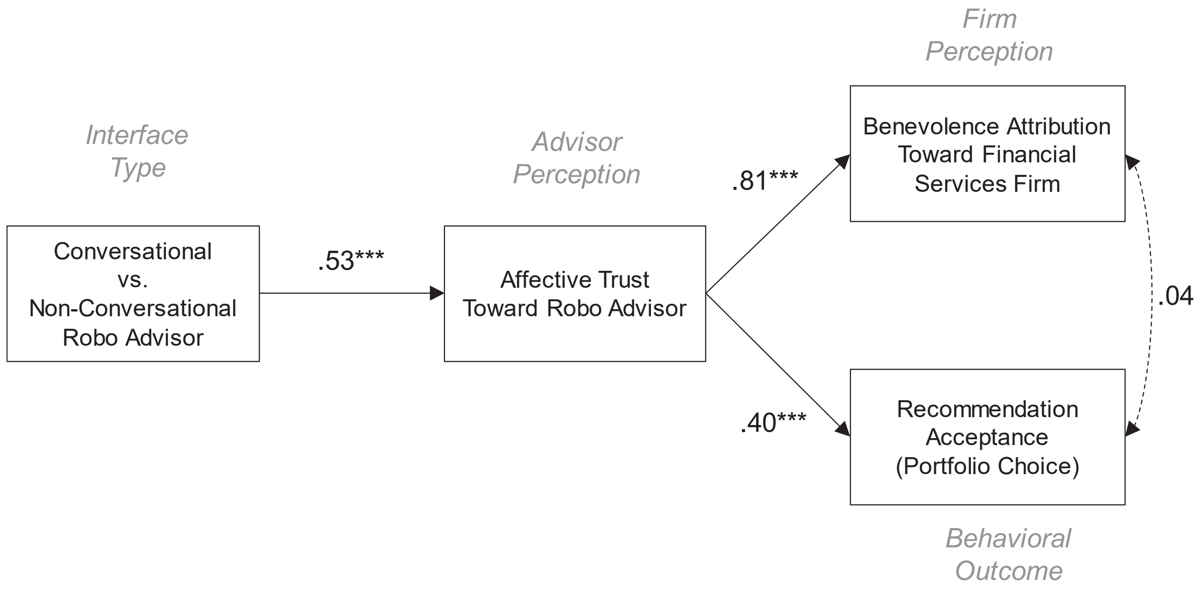
Immediately after the risk profiling task, participants completed an incentive-compatible investment game. The investment game was designed as follows: Participants received a virtual currency of 100,000 coins and had to decide on the amount they are willing to invest in the portfolio they received during the risk profiling task. Participants were informed that all asset classes may fluctuate and can produce negative and positive returns while non-invested coins will receive a return of zero. Participants then decided which percentage they wish to invest and have managed by their respective robo advisor (dependent on condition). Participants' decision was consequential, as one participant was chosen at random and received 1% of the surplus from the investment game (e.g., if a participant made a surplus 8,000 coins, they had the chance to win $80.00). The key dependent variable in this study was participants' selected amount to invest and be managed by the robo advisor. Immediately after completing the investment game, participants were forwarded to a follow-up questionnaire to assess their perceptions of affective trust toward the robo advisor ($\alpha_{TrustAff}$ = .85) and attributions of benevolence toward the financial services firm using the same items as in Study 1 ($\alpha_{FirmBenevolence}$ = .87).

*Results*

*Investment amount.* Participants in the conversational robo advisory interface condition were willing to invest a significantly larger amount managed by the robo advisor compared to participants in the non-conversational robo advisor condition ($M_{Conversational}$ = 62.71%, $M_{NonConversational}$ = 55.69%, $t$ = 2.203, $p$ < .05). Comparing the proportion of those participants who were willing to invest their entire endowment (i.e., 100%) to be managed by the robo advisor further revealed a significantly larger proportion in the conversational compared to the non-conversational robo advisor condition ($P_{Conversational}$ = 19.4%, $P_{NonConversational}$ = 7.8%; $\chi^2(1, N = 309) = 7.834$, $p$ < .01).

*Affective trust and benevolence*. Participants in the conversational robo advisory interface also attributed significantly greater levels of affective trust compared to the non-conversational robo advisory interface ($M_{Conversational}$ = 5.49, $M_{NonConversational}$ = 4.79, $t$ = 4.860, $p$ < .001). Furthermore, consumers also attributed greater benevolence toward the financial services firm in the conversational as opposed to the non-conversational robo advisor condition ($M_{Conversational}$ = 5.16, $M_{NonConversational}$ = 4.79, $t$ = 2.792, $p$ < .01). Consistent with our previous findings, the effect on attributions of benevolence was significantly mediated via perceptions of affective trust (95% CI of indirect effect [.30; 75]), rendering the residual direct effect of the advisory interface condition on benevolence non-significant ($\beta_{Conversational}$ = -.15, $t$ = 1.847, $p$ > .07).

*Path model.* Next, we estimated a path model using the lavaan package in R with robust standard errors to assess the effect of interface condition (coding the non-conversational robo advisor as 0 and the conversational robo advisor condition as 1) on attributions of benevolence and investment amount via affective trust. The path model results are summarized in Figure 3 and demonstrate that the increase in affective trust evoked by the conversational robo advisor ($\beta_{Conversational}$ = .53, $z$ = 4.873, $p$ < .001), in turn led to a significant increase in attributions of benevolence ($\beta_{TrustAff}$ = .81, $z$ = 21.773, $p$ < .001) and a significant increase in investment amount ($\beta_{TrustAff}$ = .40, $z$ = 8.076, $p$ < .001).




*Note*: *** = $p < .001$, ** = $p < .01$, * = $p < .05$; dashed lines indicate covariance estimates

**Figure 4-3:** Path model (Study 2)

*Discussion*

Study 2 demonstrates that conversational as opposed to non-conversational interfaces affect both investor perception (attributions of affective trust toward the robo advisor and benevolence of a financial services firm) and investor behavior (asset allocation toward the conversational as opposed to non-conversational robo advisor in an incentive-compatible investment game). Extending the findings of Study 1, the current study also highlights the central role of affective trust in the robo advisor and how greater levels of trust in turn affect investors' decision making using an incentive-compatible measure of behavior.

**Study 3**

The findings of Study 1 and 2 reveal initial insight on how conversational robo advisors affect consumer financial decision making, demonstrating the positive effects of conversational robo advisors in terms of a more benevolent firm perception, an increase in asset allocation toward robo advisors, and overall positive consumer sentiment. The next two studies explore further downstream consequences of these effects for consumers. Specifically, Study 3 tests whether consumers accept a portfolio recommendation even if the recommendation might be objectively wrong. That is, the current study tests whether receiving a recommendation that is inconsistent with consumers' actual risk profile varies as a function of the type of robo advisor. We focus on the acceptance of portfolio recommendations that are objectively wrong (i.e., inconsistent with investors' actual risk profile) as the more interesting case, as it provides the opportunity to directly assess whether consumers are adequately attuned to the investment advice they receive in a conversational versus non-conversational interface.

*Design and procedure*

Mirroring the experimental paradigm used in Studies 1 and 2, participants were randomly assigned to a non-conversational robo advisor condition or a conversational robo advisor condition with social cues. Participants first completed exactly the same risk assessment task as in all preceding studies. At the end of the risk assessment task, participants received a personalized portfolio recommendation based on their risk profile as in Study 2. However, this recommended portfolio was *opposite* to their actual risk profile. Specifically, if an investor's risk profile was above the midpoint of the risk profile index (risk-seeking investor), s/he received a recommended capital-preserving portfolio. On the other hand, if an investor's risk profile was below or equal to the midpoint of the risk profile index (risk-averse investor), s/he received an aggressive, growth-oriented portfolio. That is, more risk-averse investors received a portfolio recommendation with a greater percentage of stocks as opposed to bonds or money market funds, whereas risk-seeking investors received a portfolio recommendation involving a greater percentage of bonds and money market funds as opposed to stocks (see Web Appendix A1.2 for details of the matching algorithm). All other procedures were identical as in the preceding studies and we used the same scale items to assess the level of affective trust toward the robo advisor ($\alpha_{TrustAff}$ = .94) and attributions of benevolence toward the financial services firm ($\alpha_{FirmBenevolence}$ = .93). A total of 154 private investors ($M_{Age}$ = 34.57, $SD_{Age}$ = 10.62, 33.7% females) participated in this study and were preselected based on the same pre-screening criteria as before (active investors with an interest in financial advisory).

*Results*

*Affective trust and benevolence.* Participants in the conversational robo advisory interface displayed significantly greater levels of affective trust compared to the non-conversational robo advisory interface ($M_{Conversational} = 4.78$, $M_{NonConversational} = 3.24$, $t = 5.662$, $p < .001$). Furthermore, consumers attributed greater benevolence toward the financial services firm in the conversational as opposed to the non-conversational robo advisor condition ($M_{Conversational} = 4.39$, $M_{NonConversational} = 3.66$, $t = 2.916$, $p < .01$). Consistent with our previous findings, the effect on attributions of benevolence was significantly mediated via perceptions of affective trust (95% CI of indirect effect [.62; 1.47]), rendering the residual direct effect of the advisory interface condition on benevolence non-significant ($\beta_{Conversational} = -.29$, $t = 1.505$, $p > .13$).

*Recommendation acceptance.* Next, we estimated a logit model to assess the effect of the interface condition on recommendation acceptance. The findings demonstrate that participants in the conversational robo advisor condition were significantly more likely to accept the objectively incorrect portfolio recommendation compared to participants in the non-conversational robo advisor condition ($P_{Conversational} = 73.4\%$, $P_{NonConversational} = 40.0\%$; $\beta_{Conversational} = 1.42$, $z = 4.096$, $p < .001$). Critically, this effect was robust across both risk-averse investors ($P_{Conversational} = 71.2\%$, $P_{NonConversational} = 38.0\%$, $\beta_{Conversational} = 1.39$, $z = 3.406$, $p < .001$) and risk-seeking investors ($P_{Conversational} = 80.0\%$, $P_{NonConversational} = 44.0\%$, $\beta_{Conversational} = 1.42$, $z = 2.362$, $p < .05$) (see Figure 4).

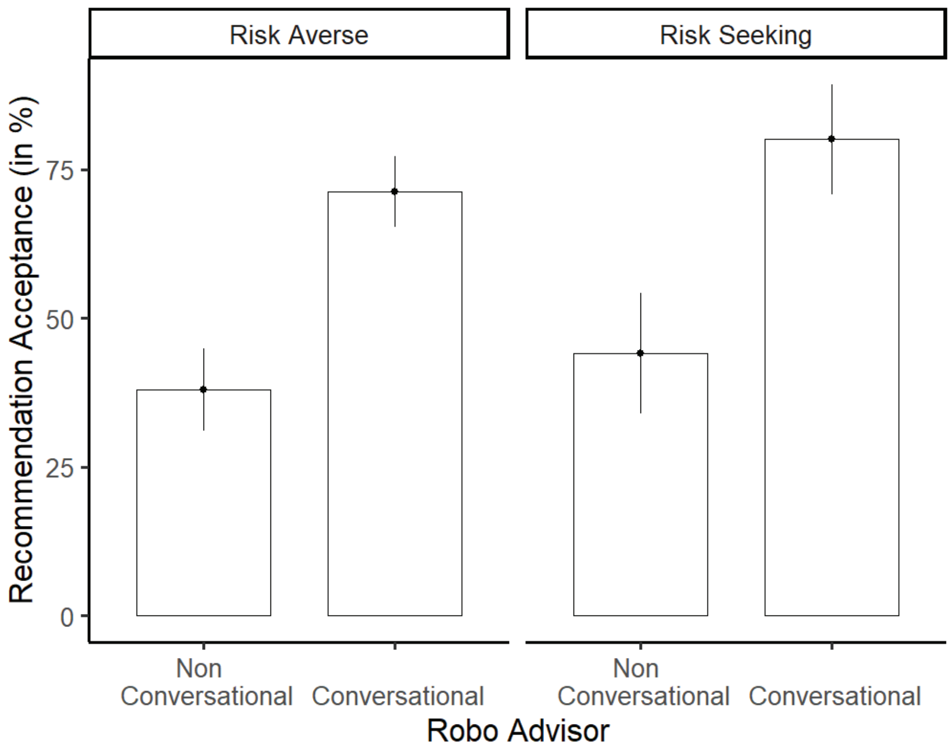
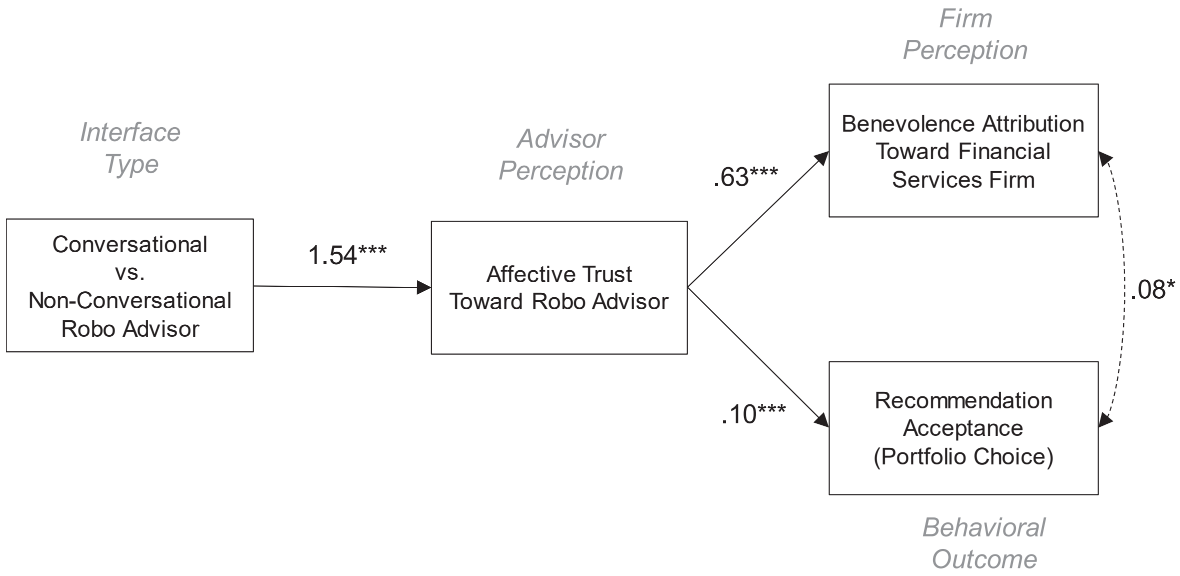



**Figure 4-4:** Recommendation acceptance by risk profile

*Path model*. Mirroring the analyses of Study 2, we estimated a path model using the lavaan package in R with robust standard errors to model the effect of the interface condition on affective trust, and how greater levels of affective trust in turn affect firm perception and recommendation acceptance. As summarized in Figure 6, we find that the pathway of the conversational interface increasing affective levels of trust ($\beta_{Conversational} = 1.54$, $z = 5.699$, $p < .001$), which in turn had a systematic influence on both firm perception and actual behavior. Specifically, affective trust in the robo advisor led to an increase in attributions of benevolence toward the financial services firm ($\beta_{TrustAff} = .625$, $z = 13.379$, $p < .001$) and greater recommendation acceptance ($\beta_{TrustAff} = .10$, $z = 5.343$, $p < .001$). These model results were robust even after controlling for differences in investors' age, risk-profile index, current level of debt, or the volume of assets that investors possessed as indicated by higher information criteria compared to our main model ($BIC_{MainModel}$: 1297.5; $BIC_{ControlModel}$: 1311.9).




*Note*: *** = $p < .001$, ** = $p < .01$, * = $p < .05$; dashed lines indicate covariance estimates.

**Figure 4-5:** Path model (Study 3)

*Discussion*

Study 3 provides evidence that greater perceptions of affective trust not only affect attributions of benevolence toward the firm but also the likelihood to accept a recommended portfolio, even if this portfolio is inconsistent with consumers' actual risk profile. Given that all participants were active private investors and that the effect was robust across both risk-averse and risk-seeking investors, these findings indicate the malleability of investor decision making in the context of conversational robo advisors and the central role of greater affective trust to alter financial decision making. We provide a more extensive discussion on the potential policy implications for consumer welfare and financial regulation in the General Discussion section.

**Study 4**

Regulations in the financial sector have led to a requirement for disclaimers during the advisory process in which (human) advisors have to inform their clients about the cost implications of their financial choices (Nussbaumer et al. 2012). Recent debates on the regulation of the robo advisor industry have discussed the necessity to hold robo advisors to the same regulatory requirements as human advisors (Baker and Dellaert 2018b). The current study is specifically designed to test whether such informational disclaimers are sufficient to debias investors and to reduce the observed over-reliance on recommendations received by a conversational robo advisor. The key hypothesis in Study 4 was to test whether this observed over-reliance to follow the algorithmic advice of a robo advisor could be attenuated by providing disclaimers that a financial services firm (and the robo advisor they employ) might act in the interest of the firm providing the advisory service.

*Design and procedure*

A total of 259 active private investors (same selection criteria as in the preceding studies) participated in this study ($M_{Age} = 34.42$, $SD_{Age} = 10.56$, 44% females). Participants were randomly assigned to a 2 (type of robo advisor: conversational vs. non-conversational) × 2 (disclaimer: present vs. absent) between-subject experiment. Participants completed the same risk assessment procedure as in the preceding studies using either a non-conversational robo advisor or conversational robo advisor with social cues. Next, all participants received a portfolio recommendation consistent with their risk profile. This portfolio comprised both actively and passively managed positions in a portfolio. All participants were led to believe that the ideal composition of a portfolio would be a spread of 30% in actively managed positions and 70% in passively managed positions (this spread was chosen based on prior work on the long-term performance of active to passive investments; see Sorensen et al. 1998). In both conditions, the advisory system recommended to increase the percentage of actively managed positions to outperform the market, even though this adjustment would cause greater annual fees. This recommendation was identical between conditions and no further information was provided about the specific range of recommended active positions. In the disclaimer-present condition, participants were then shown a prominent disclaimer prior to the opportunity to adjust the percentage of actively managed positions indicating that "The advice received on this website might be based on partial information about your actual financial situation, needs, or objectives. An automated financial advisory system might be specifically programmed to act in the interest of the company providing the advisory service." In the disclaimer-absent conditions, no additional information was provided before participants selected their preferred percentage of actively managed investments. After that, participants were free to modify the specific amount of active to passive positions in their portfolio as they wished. Participants were then forwarded to a final questionnaire and indicated their perceptions of affective trust ($\alpha_{TrustAff}$ = .92) and attributions of benevolence toward the financial services firm ($\alpha_{FirmBenevolence}$ = .93). The deviation from the initially provided 30% benchmark of actively managed funds served as the main dependent variable in this study.

*Results*

*Recommendation acceptance.* As shown in Figure 7, a two-way ANOVA with the deviation from the 30% benchmark of actively managed positions as the dependent variable and robo advisor condition (conversational vs. non-conversational) and the disclaimer condition (present vs. absent) as independent variables, revealed a significant main effect of the robo advisor condition ($F(1, 255) = 23.753$, $p < .001$), while the main effect of the disclaimer condition ($F(1, 255) = 1.405$, $p > .23$) and the interaction between both experimental factors were non-significant ($F(1, 255) = 0.935$, $p > .33$). Follow-up contrasts with Holm correction for family-wise errors revealed that participants in the conversational robo advisory condition with disclaimer ($M_{Conversational\_DisclPresent}$ = 39.73%) still deviated significantly from the 30% benchmark compared to the non-conversational robo advisor without disclaimer (vs. $M_{NonConversational\_DisclAbsent}$ = 33.04%, $t = 2.499$, $p < .05$) or with disclaimer (vs. $M_{NonConversational\_DisclPresent}$ = 32.46%, $t = 2.681$, $p < .05$). Also, the conversational robo advisory condition without disclaimer ($M_{Conversational\_DisclAbsent}$ = 43.95%) deviated significantly from both the non-conversational robo advisor condition with ($t = 4.337$, $p < .001$) or without disclaimer ($t = 4.173$, $p < .001$). Contrary to our expectation, the difference between both conversational robo advisor conditions was non-significant ($p > .26$). While we observed only minor deviations from the provided benchmark in the non-conversational robo advisor conditions (tested against zero: $M_{NonConversational}$ = 2.76%, $t = 2.791$, $p < .01$), participants in the conversational robo advisory conditions deviated significantly from the 30% benchmark (tested against zero: $M_{Conversational}$ = 11.93%, $t = 7.05$, $p < .001$).

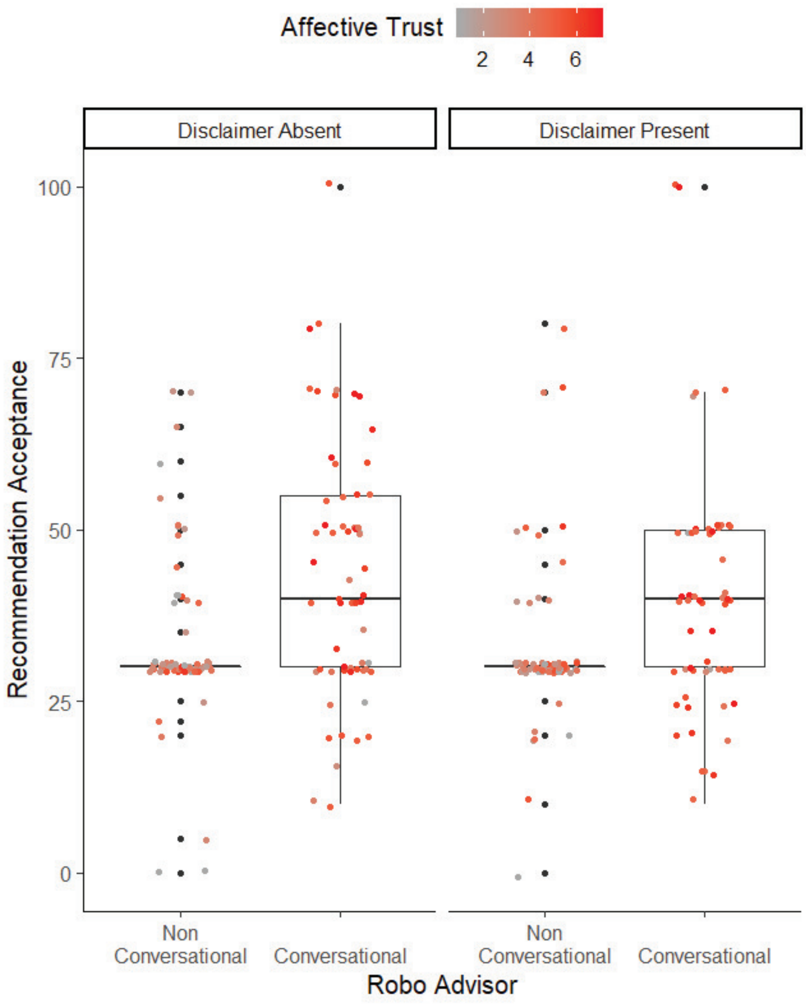
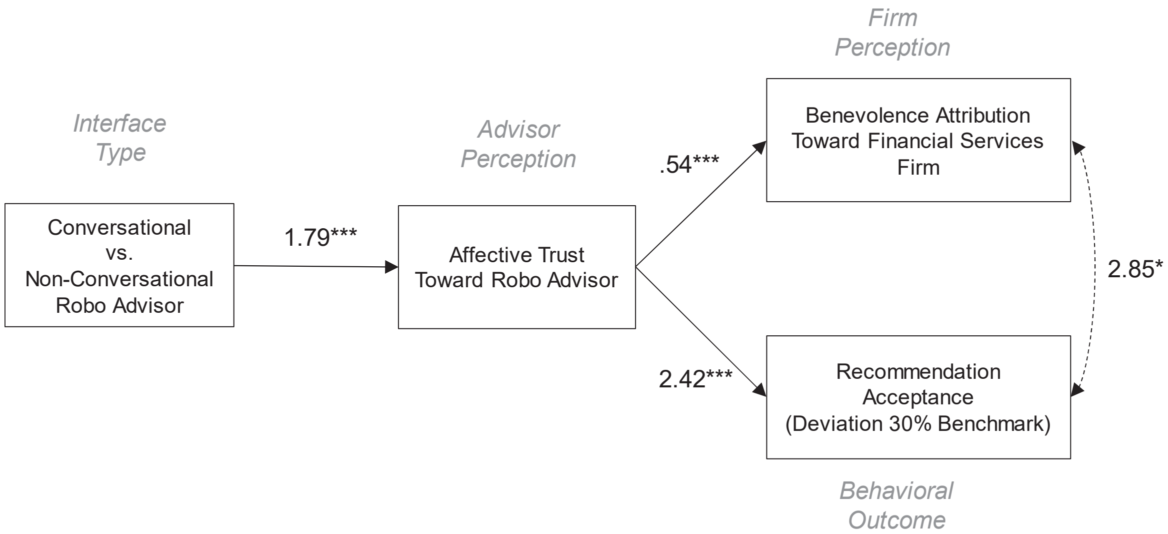



**Figure 4-6:** Deviation from 30% benchmark by interface and disclaimer conditions

*Path model.* Next, we estimated a path model using the lavaan package in R with robust standard errors and a linear link function to model the effect of the interface condition on affective trust, and how greater levels of affective trust in turn affect firm perception and recommendation acceptance in terms of deviation from the 30% benchmark. The positive effect of the conversational interface condition on affective levels of trust ($\beta_{\text{Conversational}} = 1.79$, $z = 9.913$, $p < .001$) in turn had a systematic influence on both firm perception and behavior (see Figure 8). Specifically, affective trust in the robo advisor led both to an increase in attributions of benevolence toward the financial services firm ($\beta_{\text{TrustAff}} = .54$, $z = 12.043$, $p < .001$) as well as increasing recommendation acceptance in terms of benchmark deviation ($\beta_{\text{TrustAff}} = 2.42$, $z = 4.197$, $p < .001$). These model results were robust even after controlling for differences in investors' age, risk-profile index, current level of debt, or the volume of assets that investors possessed compared to our main model ($BIC_{MainModel}$: 3893.4; $BIC_{ControlModel}$: 3904.9).




*Note*: *** = $p < .001$, ** = $p < .01$, * = $p < .05$; dashed lines indicate covariance estimates.

**Figure 4-7:** Path model (Study 4)

*Discussion*

The combined evidence of Studies 3 and 4 suggests that consumers are significantly more likely to follow investment advice from a conversational robo advisor compared to non-conversational robo advisors even if this investment advice is inconsistent with their actual risk profile (Study 3) or invokes larger annual management fees (Study 4).

## General discussion

*Theoretical implications*

The current work makes four novel contributions. First, to the best of our knowledge, the current research is the first that contrasts the effects of non-conversational robo advisors compared to conversational, dialogue-based robo advisors and how they systematically affect firm perception and investor behavior. We provide a novel conceptualization building on the psychology of human-to-human conversations and demonstrate how the conversational capacity of conversational robo advisors can systematically alter consumer financial decision making and firm perceptions. The current findings demonstrate a more positively-valenced consumer experience (see qualitative insights of Study 1), a more positive evaluation of a financial services firm, and greater recommendation acceptance (even for objectively wrong portfolio advice as in Study 3). The underlying psychological process revealed in the current research provides evidence that the affective relationship between a human investor and a robo advisor is central to our understanding of the perceptual and behavioral changes in response to interacting with a conversational robo advisor.

Second, the findings of this research contribute to prior work on trust formation between machines and humans (Coeckelbergh 2012; Hancock et al. 2011; Laursen 2013) and the role of trust as a critical component in market exchange processes (Bart et al. 2005; Hoffman, Novak, and Peralta 1999; Palmatier et al. 2012). While prior work on trust in robots primarily explored trust formation processes between physical and typically anthropomorphic robots (Laursen 2013; Wright, Sanders, and Hancock 2013), the current research provides evidence that the formation of trust is not limited to humanoid robots possessing anthropomorphic characteristics and can be induced by altering the interaction modality between humans and machines (as in the current research involving a dialogue or chat-based response modality). These findings indicate that trust formation processes and affective relationships between machines and humans may form independently of anthropomorphic traits or morphology (as with physical robots for example) and can be induced merely by altering the properties of the interface designed to structure the interaction between humans and machines.

Third, the current findings also contribute to prior work on affective trust more specifically. The formation of affective trust has been considered a long-term process (Rousseau et al. 1998), one that requires significant time evaluating the other entity to build and enhance affective levels of trust during the relationship formation phase (McAllister, Lewicki, and Chaturvedi 2006). However, the current findings demonstrate that affective levels of trust can be enhanced by changing the modality of interaction (such as a greater extent of turn-taking) or through the specific use of social cues during the interaction (such as the use of a more affect-rich language or emoticons). Thus, the current findings reveal that affective trust can be induced also more short-term, altered through the specific use of social cues, enhancing and positively affecting consumers' onboarding and investment experience.

Fourth, the current work contributes to the emerging work on immersive consumer experiences in marketing (Agarwal and Karahanna 2000; Brakus, Schmitt, and Zarantonello 2009; Schmitt, Brakus, and Zarantonello 2015) and the formation of consumer–object relationships through the use of new technologies (van Doorn et al. 2017; Hildebrand, Efthymiou, and Busquet 2020; Hoffman and Novak 2018; Melumad et al. 2020; Mende et al. 2019). Even though the time to complete the advisory process was almost twice as long in the conversational compared to the non-conversational robo advisory interface across studies (pooled duration data: $M_{Conversational}$= 288sec, $M_{NonConversational}$= 162sec, $t = 8.647$, $p < .001$), the current findings illustrate that consumers experienced the advisory process as more engaging, intimate, and enjoyable (see also the overall positive consumer sentiment revealed in the qualitative insights of Study 1). These findings highlight that the time to complete a task is a shortsighted measure compared to the importance of the actual task experience. Thus, the conversational capacity of interfaces and the effects reported in the current work illustrate how the conversational modality of a robo advisory system provides more engaging user experiences and a more enjoyable digital onboarding experience during the customer acquisition phase, positively affecting a series of perceptual and behavioral outcomes for both consumers (more positively valenced advisory experience) and the firm (more benevolent view of the firm, more positive consumer sentiment, greater willingness to follow investment advice, and overall greater trust toward a robo advisor).

*Future research*

We hope that the current research might inspire future work in two notable ways. First, future research may further examine the extent to which conversational interfaces can be specifically designed to express pre-designed "personality traits" that are consistent with the personality of the brand or firm. We see a variety of interesting and unanswered questions in the marketing and branding domain such as whether specific brand or firm associations could be expressed by tailoring the language, appearance, or other design features of a conversational interface. Second, we see great potential for future work at the marketing and finance intersection and how robo advisors might alter consumer–firm relationships more broadly. The current work hints at the possibility that the creation of more affective relationships between humans and machines might not be limited to interfaces that are intentionally anthropomorphized, but that they could be "designed" by altering the modality of interaction within the interface itself (such as building on the capacity to take turns with a richer set of social cues to induce greater affective trust). Recent research by Hoffman and Novak (2018) could provide the conceptual foundations along which future work may organize and study such emerging consumer–object and consumer–firm relationships for the effective design of conversational robo advisors. Future work may also test these effects under realistic market conditions with more consequential outcomes or further assessing the robustness of the current findings across firm types and sales channels (see also Lourenço, Dellaert, and Donkers 2020).

*Managerial and policy implications*

The findings of this research have important managerial and policy implications. First, we see great potential for the financial industry and other industries to improve the digital customer experience during the onboarding process. Repeated dyadic interaction during service encounters critically contributes to greater customer satisfaction (Solomon et al. 1985) and recent surveys suggest that private investors strongly anchor on the affective experience during these first financial advisor encounters (Darwish 2006). The turn-taking paradigm of conversational robo interfaces as illustrated in the current research demonstrates that conversational interfaces can contribute to such positive, more affective onboarding and firm experiences. Although the current research did not explicitly investigate the entire onboarding process, both future academic research and practitioners may further explore whether the trust and benevolence attributions shown in the current research translate into a more successful onboarding and customer acquisition process.

Second, we see great potential to link conversational interfaces to contextual factors during the investment selection process. For example, conversational robo advisors provide a natural interface to acquire consumers' personal interests either explicitly through direct interaction or unobtrusively by using referrer or cookie information from the browser. This information can be used to tailor specific recommendations throughout the advisory process, similar to prior work in retargeting and journey analytics on consumer websites (Hildebrand and Schlager 2019; Urban et al. 2014). These avenues hold the potential to provide a more personalized customer experience based on information that is generally unobservable in traditional human-to-human service encounters. Thus, conversational interfaces provide a powerful tool to shape and unlock value in digital customer journeys and provide a more tailored one-to-one customer experience.

Finally, the current findings also address recent debates on the increasing share of actively managed funds among robo advisor providers. For example, Betterment's "smart beta" strategy was introduced recently for those investors who wish to "outperform the market" (Salmon 2018). Similarly, companies such as Wealthfront have strayed from putting investors solely into a low-cost indexing strategy and decided to invest 20% of their clients' assets into an internal "risk parity" fund that invested in derivatives such as total return swaps. Both cases caused significant negative press in the media and among investors, questioning these changes in the firm's investment strategy (Salmon 2018). The findings of Study 4 suggest that conversational robo advisors can lead to an increase in the acceptance of a greater number of actively managed funds in consumers' portfolio without jeopardizing perceptions of trust or leading to negative perceptions of the firm. However, the combined evidence of Studies 3 and 4 also underline recent calls for missing regulations of robo advisors in the financial industry (Baker and Dellaert 2018b) and that consumers are not sufficiently attuned to question the received automated investment advice.

## Conclusion

As robo advisors and other digital advisory systems grow in scale and modalities (from traditional, non-conversational interfaces to the natural language processing capabilities of conversational interfaces), a new cross-disciplinary science of new technologies and augmented consumer decision making is emerging. We hope that the current work inspires more work in this emerging field on the consumer psychology of new technologies and the implications for consumer welfare in the digital economy.

## References

Adam, Martin, Jonas Toutaoui, Nicolas Pfeuffer, and Oliver Hinz (2019), "Investment decisions with robo advisors: the role of anthropomorphism and personalized anchors," in *Proceedings of the 27th European Conference on Information Systems (ECIS)*, 0–18.

Agarwal, Ritu, and Elena Karahanna (2000), "Time Flies When You're Having Fun: Cognitive Absorption and Beliefs about Information Technology Usage," *MIS Quarterly*, 24(4), 665.

Andrus, Danielle (2014), "3 Reasons Robo-Advisors Are Taking Off," *Investment Advisor Magazine*, 34(12), 5–5.

Araujo, Theo (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, 183–89.

Aronson, Elliot (1999), "The Power of Self-Persuasion," *American Psychologist*, 54(11), 875–84.

Baker, Tom (2018), "Regulating Robo Advice Across the Financial Services Industry Regulating Robo Advice Across the," *Penn Law: Legal Scholarship Repository*.

Baker, Tom, and Benedict Dellaert (2018), "Regulating robo advice across the financial services industry," *Iowa Law Review*, 103(2), 713–50.

Barrons (2017), "As Robo-Advisors Cross $200 Billion in Assets, Schwab Leads in Performance," <https://www.barrons.com/articles/as-robo-advisors-cross-200-billion-in-assets-schwab-leads-in-performance-1517509393>.

Bart, Yakov, Venkatesh Shankar, Fareena Sultan, and Glen L. Urban (2005), "Are the Drivers and Role of Online Trust the Same for All Web Sites and Consumers? A Large-Scale Exploratory Empirical Study," *Journal of Marketing*, 69(4), 133–52.

BenMark, Gadi, and Dilip Venkatachari (2016), "Messaging Apps Are Changing How Companies Talk with Customers," *Harvard Business Review*.

Betancourt, Hector (1990), "An Attribution-Empathy Model of Helping Behavior," *Personality and Social Psychology Bulletin*, 16(3), 573–91.

Bickmore, Timothy, and Justine Cassell (2000), "' How about this weather ?' Social Dialogue with Embodied Conversational Agents REA : An Embodied Conversational Real-," (Clark 1996).

——— (2001), "Relational agents: A Model and Implementation of Building User Trust," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, Seattle, WA, 396–403.

Brakus, J. Joško, Bernd H Schmitt, and Lia Zarantonello (2009), "Brand Experience: What Is It? How Is It Measured? Does It Affect Loyalty?," *Journal of Marketing*, 73(3), 52–68.

Cappela, Joseph N. (1985), "Production Principles for Turn-Taking Rules in Social Interaction: Socially anxious vs. socially secure persons," *Journal of Language and Social Psychology*, 4(3 & 4).

Cassell, Justine, and Kristinn R. Thórisson (1999), "The power of a nod and a glance: envelope vs. emotional feedback in animated conversational agents," *Applied Artificial Intelligence*, 13(4–5), 519–38.

Cho, Jinsook (2006), "The mechanism of trust and distrust formation and their relational outcomes," *Journal of Retailing*, 82(1), 25–35.

Coeckelbergh, Mark (2012), "Can we trust robots?," *Ethics and Information Technology*, 14(1), 53–60.

Dale, Robert (2016), "The Return of the Chatbots," *Natural Language Engineering*, 22(5), 811–17.

Darwish, Ryan (2006), "Trust, Authenticity, and the Advisor," *Journal of Financial Planning*, 19(6), 32–35.

Davenport, Thomas, Abhijit Guha, Dhruv Grewal, and Timna Bressgott (2020), "How artificial intelligence will change the future of marketing," *Journal of the Academy of Marketing Science*, 48(1), 24–42.

Day, Min-yuh, Jian-ting Lin, and Yuan-chih Chen (2018), "Artificial Intelligence for Conversational Robo-Advisor," *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 1057–64.

van Doorn, Jenny, Martin Mende, Stephanie M. Noble, John Hulland, Amy L. Ostrom, Dhruv Grewal, and J. Andrew Petersen (2017), "Domo Arigato Mr. Roboto," *Journal of Service Research*, 20(1), 43–58.

Faloon, Michael, and Bernd Scherer (2017), "Individualisation of Robo Advice," *Journal of Wealth Management*, 20(1), 30–36.

Fiske, Susan T., Amy J.C. Cuddy, and Peter Glick (2007), "Universal Dimensions of Social Cognition: Warmth and Competence," *Trends in Cognitive Sciences*, 11(2), 77–83.

Fransen, Marieke L., Edith G. Smit, and Peeter W. J. Verlegh (2015), "Strategies and motives for resistance to persuasion: an integrative framework," *Frontiers in Psychology*, 6.

Gomber, Peter, Jascha-Alexander Koch, and Michael Siering (2017), "Digital Finance and FinTech: Current Research and Future Research Directions," *Journal of Business Economics*, 87(5), 537–80.

Grice, H.P. (1975), *Logic and conversation*, *Syntax and Semantics, Volume 3: Speech Acts*, (P. Cole and J. L. Morgan, eds.), Academic Press.

Hancock, Peter A., Deborah R. Billings, Kristin E. Schaefer, Jessie Y.C. Chen, Ewart J. De Visser, and Raja Parasuraman (2011), "A meta-analysis of factors affecting trust in human-robot interaction," *Human Factors*, 53(5), 517–27.

He, Hongwei, Yan Li, and Lloyd Harris (2012), "Social identity perspective on brand loyalty," *Journal of Business Research*.

Hexmoor, Henry, Shahram Rahimi, and Rachil Chandran (2008), "Delegations guided by trust and autonomy," *Web Intelligence and Agent Systems*, 6(2), 137–55.

Hoffman, Donna L., and Thomas P. Novak (2018), "Consumer and Object Experience in the Internet of Things: An Assemblage Theory Approach," *Journal of Consumer Research*, 44(6), 1178–1204.

Hoffman, Donna L., Thomas P. Novak, and Marcos Peralta (1999), "Building consumer trust online," *Communications of the ACM*, 42(4), 80–85.

Hohenberger, Christoph, Chaiwoo Lee, and Joseph F. Coughlin (2019), "Acceptance of robo-advisors: Effects of financial experience, affective reactions, and self-enhancement motives," *Financial Planning Review*, 2(2).

Johnson, Devon, and Kent Grayson (2005), "Cognitive and Affective Trust in Service Relationships," *Journal of Business Research*, 58(4), 500–507.

Jones, Gareth R., and Jennifer M. George (1998), "The experience and evolution of trust: Implications for cooperation and teamwork," *Academy of Management Review*.

Jung, Dominik, Verena Dorner, Florian Glaser, and Stefan Morana (2018), "Robo-Advisory: Digitalization and Automation of Financial Advisory," *Business and Information Systems Engineering*, 60(1), 81–86.

Jung, Dominik, Verena Dorner, Christof Weinhardt, and Hakan Pusmaz (2017), "Designing a Robo-Advisor for Risk-Averse, Low-Budget Consumers," *Electronic Markets*, 1–14.

Kilic, Mehmet, Peter Heinrich, and Gerhard Schwabe (2015), "Coercing into Completeness in Financial Advisory Service Encounters," *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing - CSCW '15*, (March), 1324–35.

Köhler, Clemens F, Andrew J Rohm, Ko de Ruyter, and Martin Wetzels (2011), "Return on Interactivity: The Impact of Online Agents on Newcomer Adjustment," *Journal of Marketing*, 75(2), 93–108.

Laursen, Lucas (2013), "Robot to human: 'Trust me,'" *IEEE Spectrum*, 50(3), 18–18.

Levinson, Stephen C. (2016), "Turn-taking in Human Communication - Origins and Implications for Language Processing," *Trends in Cognitive Sciences*, 20(1), 6–14.

Looije, Rosemarijn, Mark A. Neerincx, and Fokie Cnossen (2010), "Persuasive Robotic Assistant for Health Self-Management of Older Adults: Design and Evaluation of Social Behaviors," *International Journal of Human Computer Studies*, 68(6), 386–97.

Lourenço, Carlos J.S., Benedict G.C. Dellaert, and Bas Donkers (2020), "Whose Algorithm Says So: The Relationships Between Type of Firm, Perceptions of Trust and Expertise, and the Acceptance of Financial Robo-Advice," *Journal of Interactive Marketing*.

Luo, J. T., Peter McGoldrick, Susan Beatty, and Kathleen A. Keeling (2006), "On-screen characters: Their design and influence on consumer trust," *Journal of Services Marketing*, 20(2), 112–24.

McAllister, Daniel J., Roy J. Lewicki, and Sankalp Chaturvedi (2006), "Trust in developing relationships: From theory to measurement," in *Academy of Management 2006 Annual Meeting: Knowledge, Action and the Public Concern, AOM 2006*.

Mende, Martin, Maura L. Scott, Jenny van Doorn, Dhruv Grewal, and Ilana Shanks (2019), "Service Robots Rising: How Humanoid Robots Influence Service Experiences and Elicit Compensatory Consumer Responses," *Journal of Marketing Research*, 56(4), 535–56.

Monti, Marco, Vittorio Pelligra, Laura Martignon, and Nathan Berg (2014), "Retail investors and financial advisors: New evidence on trust and advice taking heuristics," *Journal of Business Research*, 67(8), 1749–57.

Musto, Cataldo, Giovanni Semeraro, Pasquale Lops, Marco De Gemmis, and Georgios Lekkas (2015), "Personalized Finance Advisory Through Case-Based Recommender Systems and Diversification Strategies," *Decision Support Systems*, 77, 100–111.

Nielsen, Finn Årup (2011), "A new ANEW: Evaluation of a word list for sentiment analysis in microblogs," *CEUR Workshop Proceedings*, 718, 93–98.

Nussbaumer, Philipp, Inu Matter, Gian Reto À Porta, and Gerhard Schwabe (2012), "Designing for Cost Transparency in Investment Advisory Service Encounters," *Business and Information Systems Engineering*.

Palmatier, Robert W., Mark B. Houston, Rajiv P. Dant, and Dhruv Grewal (2012), "Relationship Velocity: Toward a Theory of Relationship Dynamics," *Journal of Marketing*.

Pástor, Ľuboš, and Robert F. Stambaugh (2002), "Investing in equity mutual funds," *Journal of Financial Economics*, 63(3), 351–80.

Pavlou, A. Paul (2003), "Consumer acceptance of electronic commerce: Integrating trust and risk with the technology acceptance model," *International Journal of Electronic Commerce*, 7(3), 69–103.

Rousseau, Denise M., Sim B. Sitkin, Ronald S. Burt, and Colin Camerer (1998), "Not so different after all: A cross-discipline view of trust," *Academy of Management Review*.

Salmon, Felix (2018), "Beware of Robo-Advisors Bearing Low Fees," *Wired*.

Schlosser, Ann E, Tiffany Barnett White, and Susan M Lloyd (2006), "Converting Web Site Visitors into Buyers: How Web Site Investment Increases Consumer Trusting Beliefs and Online Purchase Intentions," *Journal of Marketing*, 70(2), 133–48.

Schmitt, Bernd H., Joško Brakus, and Lia Zarantonello (2015), "The current state and future of brand experience," *Journal of Brand Management*, 21(9), 727–33.

Schweitzer, Vivien (2019), "Only 20% of Affluent Millennials are Using Robo-Advisors," <https://www.investopedia.com/study-affluent-millennials-are-warming-up-to-robo-advisors-4770577>.

Solomon, Michael R., Carol Surprenant, John A. Czepiel, and Evelyn G. Gutman (1985), "A Role Theory Perspective on Dyadic Interactions: The Service Encounter," *Journal of Marketing*, 49(1), 99.

Sorensen, Eric H., Keith L. Miller, and Vele Samak (1998), "Allocating between Active and Passive Management," *Financial Analysts Journal*, 54(5), 18–31.

Sprecher, Susan, Stanislav Treger, Joshua D. Wondra, Nicole Hilaire, and Kevin Wallpe (2013), "Taking Turns: Reciprocal Self-Disclosure Promotes Liking in Initial Interactions," *Journal of Experimental Social Psychology*, 49(5), 860–66.

Tedesco, Darren (2015), "I, Robo-Adviser? Creating the Blended Adviser Experience.," *Journal of Financial Planning*, 28(1), 17–19.

Thomaz, Felipe, Carolina Salge, Elena Karahanna, and John Hulland (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.

Urban, Glen L., Guilherme (Gui) Liberali, Erin MacDonald, Robert Bordley, and John R. Hauser (2014), "Morphing Banner Advertising," *Marketing Science*, 33(1), 27–46.

Wall Street Journal (2014), "Taking Stock of Automated Financial Advisers," <https://www.wsj.com/articles/taking-stock-of-automated-financial-advisers-1407519771>.

Wiemann, John M., and Mark L. Knapp (1975), "Turn-taking in Conversations," *Journal of Communication*, 25(2), 75–92.

Williams, Oscar (2017), "Banks are looking to use artificial intelligence in almost every part of their business: Here's how it can boost profits," *Business Insider*.

Wright, Julia, Tracy Sanders, and Peter A. Hancock (2013), "Identifying the role of attributions in human perceptions of robots," in *Proceedings of the Human Factors and Ergonomics Society*, 1288–92.

Xie, Yi, and Siqing Peng (2009), "How to repair customer trust after negative publicity: The roles of competence, integrity, benevolence, and forgiveness," *Psychology and Marketing*, 26(7), 572–89.

Yousafzai, Shumaila Y., Gordon Robert Foxall, and John Gordon Pallister (2010), "Explaining Internet Banking Behavior: Theory of Reasoned Action, Theory of Planned Behavior, or Technology Acceptance Model?," *Journal of Applied Social Psychology*, 40(5), 1172–1202.

# 5. ARTICLE IV

Bergner, A., Hartmann, J., and Hildebrand, C. (ready for submission). Conferring Minds to Machines: A Deep Learning Approach to Mind Perception, Smart Object Relationships, and Task Delegation. *Journal of Consumer Research.*

**Abstract**

Smart objects (such as digital voice assistants) are dramatically changing how consumers relate to technology. This research proposes a novel methodological approach to automated text analysis to identify and understand the attribution of a "mind" to smart objects and its implications on consumer behavior. Specifically, we train a state-of-the-art deep contextual language model using transfer learning to unobtrusively classify mind perception from unsolicited, consumer-generated text. We provide a rich set of interpretable linguistic features using manifold methods, demonstrating that the extent of mind perception provides novel insights into the consumer-smart object relationships and successfully predicts differences in consumers' willingness to delegate a broad set of tasks to it.

*Keywords*: Mind Perception, Theory of Mind, Digital Voice Assistants, Smart Objects, Language Models, Interpretable Machine Learning, Deep Learning, Natural Language Processing

# Conferring Minds to Machines: A Deep Learning Approach to Mind Perception, Smart Object Relationships, and Task Delegation

Perceiving a mind in others is paramount to navigating the world around us. A theory of mind, defined as the attribution and understanding of internal mental states in another entity, is the hallmark of our social cognitive skills, enabling everyday interactions and task coordination with others (Goldstein and Winner 2012). The proliferation of AI-enabled smart objects that appear and "behave" increasingly more human-like raises fundamental questions about the role and impact of mind perception in the consumer technology landscape. Recent research on smart object relationships suggests that consumers may form different types of relationships with such objects (Hoffman and Novak 2018; Novak and Hoffman 2018). While some consumers perceive such objects merely as a medium to complete a task, others tend to attribute more human mental states to them, fundamentally affecting consumer experiences and extent of usage (Schweitzer et al. 2019; Waytz et al. 2014; Yang, Aggarwal, and McGill 2020). Consider for example the following two real customer reviews:

> *Customer A:* "Alexa is a very cool home assistant. She helps me complete tasks without having to do much. I ask her to calculate math problems for me, she helps me create shopping lists, she also helps me to remember to do things at specific times because I am very forgetful. If you need an assistant without having to dish out a salary this would be great for you. She's a great addition to our whole family."

> *Customer B*: "I mainly use the Amazon Echo as a timer in the kitchen. I will set multiple timers with it as I cook. This is a very useful feature. It is simple to command and was very easy to set-up in a matter of minutes. I sometimes use it for music but not very often as I don't have a good speaker to use it with."

It is apparent that there are nuances between those two exemplary customer reviews in terms of how each customer addresses the smart speaker ("Alexa" as the assistant and "her" as opposed to "Echo" as the device and "it"), the implicit relationship they seem to have formed with the object ("a cool assistant" and "a great addition to the family" as opposed to a simple "timer" with "useful features") as well as the range of tasks they engage in with it ("helps me to complete tasks" as opposed to "I use it (…) to"). This may indicate that customer A attributes more “mind” to the assistant as opposed to customer B, who seems to view Alexa as a mere tool. Furthermore, this exemplary comparison suggests that customer reviews may hold interesting insights into the relationship consumers develop with their smart objects and the extent of tasks they delegate to them. However, while previous marketing research has successfully employed text analysis for a variety of applications, such as uncovering consumer sentiments (see Berger et al. 2020; Hartmann et al. 2019; Humphreys and Wang 2018 for reviews), no off-the-shelf dictionary (e.g., LIWC, Pennebaker et al. 2015) or alternative text classifier exists to classify mind perception terms. Furthermore, compared to traditional text classification tasks, identifying the extent of mind perception from unsolicited reviews is challenging due to the absence of a theory-based mapping of mind perception dimensions and text objects, as well as the subtlety of its signals.

Following recent calls for multi-method approaches and natural language processing (Berger et al. 2020; Schmitt et al. 2021), this research proposes a novel approach to developing theory and conducting research on unstructured, user-generated text, leveraging the power of deep contextual language models for automated text analysis and applying it to shed light on mind perception in AI-enabled smart objects. Based on a large-scale panel of smart object users, we develop a state-of-the-art deep learning text classifier, demonstrating the accurate prediction of mind perception from unsolicited, unstructured customer reviews and provide a rich set of interpretable linguistic features associated with higher mind perception, using manifold methods (LDA topic models, Lasso regressions, word frequency assessments and novel techniques from the field of interpretable machine learning). We validate the classifier by applying it to a large set of customer reviews scraped from a major online review site and finally, demonstrate its transferability to other categories of smart objects.

In what follows, we first review fundamental literature on theory of mind and argue that attributing a mind to a smart object provides novel insights into how people relate to these technologies. We then discuss the advantages of using a deep contextual language model to assess and classify mind perception unobtrusively from customer reviews and provide an in-depth discussion of our proposed approach. Next, we present the results of three studies that were designed to test and validate our theorizing and proposed method.

We conclude with a discussion on the power of deep contextual language models to detect nuanced variations and uncover novel insights in unstructured texts, opening up new methodological opportunities in consumer research.

## Theoretical Background and Hypotheses

### *Theory of Mind as a Fundamental Human Property*

Thinking whether other entities possess a mind inspired research from interpersonal perception in humans (Gray et al. 2007; Waytz, Gray, Epley and Wegner 2010) to seminal court cases on human rights for animals (Premack and Woodruff 1978) or even to theological beliefs (Barrett and Keil 1996). Even non-human entities possessing limited human-like characteristics, such as a robot moving at a human-like speed or a human-looking avatar in a chatbot (Riek et al. 2009; Araujo 2018; Bergner et al. 2019) are often sufficient to create the belief of a mind in the form of agency (defined as the ability to think and plan rationally) and a subjective experience (defined as the ability to experience conscious emotions) in non-human objects. Indeed, prior research has shown that attributions of such mental capacities to other entities tends to unfold naturally and automatically, fundamentally shaping how we perceive and relate to others (Gray et al. 2007; Haslam et al. 2008; Waytz, Gray, et al. 2010). In contrast, denying essential human capacities such as agency and emotionality leads to a sense of dehumanization, in which entities lacking such features are likened to machines and perceived as cold, passive, and shallow (Haslam et al., 2008). Hence, perceiving a mind in others is a fundamental aspect of social perception.

The attribution of a mind in another entity is a fundamental human capacity crucial to human interaction and serves the function of making inferences about another entity's behavior. Indeed, as virtually all human behavior is driven by underlying mental states, humans learn early on in childhood how to make inferences about the internal state of another being, such as its intentions, knowledge, and beliefs, in order to make sense of its actions (Birch et al., 2017). Yet, this reasoning about internal, unobservable states in other entities has been shown to extend far beyond human-to-human interaction to reasoning about our pets' behavior and even the behavior of objects such as cars or computers (Epley et al. 2007; Waytz, Cacioppo, et al. 2010). Thus, attributing a mind to another entity and reasoning about its internal mental states is one of the most fundamental human capacities, allowing us to make sense of, and connect to other entities, whether they are human, animals, or objects (Epley et al. 2008; Yang et al. 2020).

*Mind Perception and Language*

The strong connection between communication and mind perception has long been established in developmental psychology, demonstrating the existence of important links between language acquisition and the ability to reason about another person's mental state (Astington and Baird 2005). Specifically, research has shown that a child's participation in two-way conversations leads to his or her understanding of mental states in others (Bretherton et al. 1981; Dunn et al. 1991). Furthermore, in order to take the perspective of and reason about an internal, unobservable state of another entity, people need to develop a language that allows them to describe the behavior of an entity in terms of its mental state. Specifically, children's theory of mind development is directly linked to their acquisition of specific lexical terms used to refer to mental states, such as terms related to the subjective experience (e.g., "happy", "sad", "like", "love", "want"; Bretherton and Beeghly 1982) and cognitive abilities (e.g., "know", "think", "remember"; Bretherton and Beeghly 1982) of the other entity.

In the context of sophisticated technology, even the use of more simple, anthropomorphic terms has been shown to encourage the attribution of human mental states to a non-human entity. For example, in their study on autonomous vehicles, Waytz et al., (2014) show that using anthropomorphized terms such as using "she" rather than "it" when referring to the car and calling the car a name, in addition to giving it a voice in the first person, significantly increased the attribution of mental capacities to it, such as its capacity to feel what is happening, as well as its ability to anticipate actions.

Furthermore, it has been shown that people tend to automatically use such semantics when reflecting on and describing the action of other entities, even if they are non-human. Indeed, even when trying to describe the interaction of abstract objects, people refer back to attributions of mental states, in order to make sense of and be able to label the actions of these abstract objects (Abell et al. 2000). Thus, given the established importance of the link between language and theory of mind (Astington and Baird 2005), we expect to find linguistic markers of mental attribution, such as lexical terms related to the personification and mental abilities of the other entity.

*Mind Perception as a Relational Construct*

Research on dehumanization shows that denying mental capacities to others is linked to indifference and disengagement, and an increased psychological distance with the other entity (Haslam et al., 2006). Relatedly, prior research in marketing has shown that attributing human traits and intentions to brands and products leads to an increase in liking and closeness with the brand or product (Delbaere, McQuarrie, and Phillips 2011; Fournier 1998; Landwehr et al. 2011; Thomson, MacInnis, and Park 2005a). As discussed earlier, attributing mental capacities to another entity encourages perspective taking by actively thinking about the experience of the other entity, an aspect that is fundamental to effective communication between two entities (Mead 1934). Indeed, Airenti (2018) argues that, rather than a belief system, attributing mental states to another entity should be considered a means to establish a relationship with it, primarily through the medium of communication. Communicating about one's internal state enables perspective taking and allows to explain and predict future actions more accurately (Carruthers & Smith, 1996). The implications of such perspective taking for interpersonal interactions and relationships are wide, ranging from eliciting positive attributions and empathy (Batson et al. 1995; Parker and Axtell 2001) to more effective interpersonal relating (Parker, Atkins, and Axtell 2008) and active trust building (Williams 2007). Thus, we expect that higher mind perception in smart objects will lead to closer, more intimate relationships.

While Hoffman and Novak (2018) argue that this perspective taking should be based on an object-centric rather than a human-centric anthropomorphism, thus seeing the object's experience for what it really is rather than attributing human mental states to it, they admit that consumers may lack the experience and vocabulary to relate to smart objects in purely object-centric terms, and may need to employ metaphors of more human-centric emotions and mental states in order to understand object experiences in the IoT. We therefore argue that mind perception can be seen as a form of perspective-taking through communication and a means of consumers to make sense of the subjective experience of a smart object in order to determine how they should relate to it.

*Mind Perception as a Predictor of Task Delegation*

The ability of smart objects to interact with consumers and independently execute tasks makes them active partners in the consumption process (Hoffman and Novak 2018). A closer, and more intimate relationship increases the belief that one's counterpart will act in ways that are helpful rather than harmful, increasing cooperation and coordinated social interactions (Gächter et al. 2004; Rousseau et al. 1998). Relatedly, recent research shows that higher perceptions of mental capacities in robots and autonomous cars leads to an increase in the attribution of responsibilities to them during collaborative tasks (Hinds, Roberts, and Jones 2004; Waytz et al. 2014).

Novak and Hoffman (2019) suggest that consumers can take one of two primary modalities in their interaction with smart objects and either enable or restrict the object's actions within its broader assemblage of connected devices, thus either allowing or denying an object's ability to execute tasks on behalf of the consumer. Furthermore, interpersonal perception has been shown to be an important determinant of delegating decisions to another entity (Aggarwal and Mazumdar 2008; Komiak, Wang, and Benbasat 2004). Specifically, familiarity with the surrogate and relationship closeness are key determinants in selecting a decision surrogate and delegating a decision (Komiak and Benbasat 2006; Rosen and Olshavsky 1987). Thus, taken together, we expect that relationship closeness with a smart object will positively influence consumers' willingness to delegate a broader range of tasks to it.

## Method

*Automated Text Analysis*

The strong connection between language and mind perception suggests that there may be subtle cues in text that would allow us to identify different types of consumer-smart object relationships based on the way in which consumers describe a smart object. Indeed, research has shown that text can reflect important information about the text producer, such as their personality or mood (Moon and Kamakura 2017), mental states and emotions (Rude, Gortner, and Pennebaker 2004; Settanni and Marengo 2015), or intentions (Ludwig et al. 2016; Netzer, Lemaire, and Herzenstein 2019; Seraj, Blackburn, and Pennebaker 2021). Furthermore, text can provide insights into the attitude and relationship of the producer with another entity or object (Berger et al. 2020; Packard and Berger 2019). For example, Seraj et al. (2021) were able to identify markers in language that can predict relationship break-ups before they happened based on an increase in self-focus words (I-words) and a decrease in cognitive thinking words (words such as "understand" and "meaning") used in social media posts. Thus, given the important relationship between language and mind perception, we believe that automated text analysis from unsolicited customer reviews may shed light on the extent of mind perception and the underlying consumer-smart object relationships.

Automated text analysis has been a powerful tool to extract insights from unstructured sources of data, from predicting election outcomes (Bovet, Morone, and Makse 2018), to enabling rapid disaster response (Kryvasheyeu et al. 2016), and tracking public wellbeing (Hills et al. 2019; Zheng et al. 2019). Previous marketing research has successfully employed text analysis for a variety of applications (Berger et al. 2020; Hartmann et al. 2019; Humphreys and Wang 2018). These applications include automatically creating market maps (Netzer et al. 2012), extracting customer needs (Timoshenko and Hauser 2019) and understanding consumer sentiments (Ordenes et al. 2017), as well as predicting loan defaults (Netzer et al. 2019), sales conversions (X. Liu, Lee, and Srinivasan 2019) or content sharing (Ordenes et al. 2019) from text.

One of the major advantages of automated text analysis is that it releases researchers from the restraint of scripted questions and the biases introduced by self-report measures and it allows to quantify and classify behavior naturally, from the information that is available in unstructured user-generated text (Berger et al. 2020). As the attribution of mental capacities to and the underlying relationships with smart objects are nuanced, context-dependent, and arguably complex, extracting information from unsolicited customer reviews provides an unexplored methodology to gain novel and nuanced insight into consumers and the relationships they form with smart objects.

*Deep Contextual Language Models*

Sentiment analysis is among the most popular tasks in natural language processing (NLP, Hirschberg and Manning 2015), often relying on simple lexicon-based approaches to classify text data into broad categories such as "positive" versus "negative" sentiment

(Hartmann et al. 2019). While sentiments can be detected in text with various "off-the-shelf" dictionaries such as LIWC (Pennebaker et al. 2015) or the Evaluative Lexicon (Rocklage and Fazio 2020), to the extent of our knowledge, no such handcrafted lexicon or alternative text classifier exists to unobtrusively classify the extent of mind perception in unstructured, user-generated text. Compared to traditional text classification tasks such as sentiment analysis, where accuracy levels across methods on average exceed 80% (Heitmann et al. 2020), identifying and classifying mind attribution to machines from unsolicited customer reviews is challenging due to the absence of a theory-based mapping of mind perception dimensions and linguistic features (i.e., tokens or terms).

Recent developments in NLP however provide a novel approach to extract subtle cues from unstructured text. Introduced in late 2018, BERT, which stands for Bidirectional Encoder Representations from Transformers (Devlin et al. 2019; Vaswani et al. 2017) and its successor RoBERTa (Y. Liu et al. 2019) represented a major break-through in NLP. BERT and RoBERTa are deep neural networks specialized on NLP tasks that are pretrained in a self-supervised manner on massive open-domain text corpora (e.g., 13 GB of text data for BERT and over 160 GB for RoBERTa) by predicting a missing word in a sequence, so-called masked language modeling. This pretraining procedure allows these large-scale *deep contextual language models* to learn semantic and syntactic patterns of human language without any explicit supervision signals (Manning et al. 2020). The amount of data and number of parameters (i.e., 125 million parameters for the base architecture of RoBERTa) make this self-supervised pretraining costly. However, once a language model has been trained, it can be fine-tuned to a variety of other tasks outside of the domain and data on which the model has been trained. This process, known as transfer learning, requires only minor adaptations of the deep neural network's architecture and labeled training data that is by multiple orders of magnitude smaller compared to training a model from scratch (Howard and Ruder 2018). The main advantage of deep contextual language models over established embedding approaches such as word2vec (Mikolov et al. 2013) is that they can represent words based on the context in which they appear (Y. Liu et al. 2019). Consider, for example, the word "duck" in the following two sentences: (a) "He is cooking *duck* for his whole family*"* vs. (b) "*Duck* down before you get hit!" Unlike word2vec, embedding the word "duck" using RoBERTa yields different representations in latent space for these two sentences. Unsurprisingly, Transformer-based language models have outperformed extant NLP methods across a large variety of tasks and are consistently ranked at the top of popular benchmarks such as the General Language Understanding Evaluation (GLUE) benchmark.

However, the downside of using deep learning, and specifically deep contextual language models such as RoBERTa, is that they are often considered a "black box", sacrificing interpretability for prediction accuracy (Rai 2020). Indeed, a model such as RoBERTa learns associations between nuanced linguistic features and mind perception by itself and is made up of highly non-linear associations due to the complex connections between the underlying layers of the neural network. In contrast to linear regression models, whose parameters can be interpreted intuitively as linear relationships between the dependent in independent variables, the inner workings of deep contextual language models are difficult to interpret directly.

The present research provides a novel approach, combining deep learning with more traditional text mining methods and leveraging post-hoc interpretability techniques approximating the deep learning model with simpler models in order to provide an explanation for its predictions. We believe that this novel approach opens up an exciting new avenue for consumer research on unstructured text by providing higher prediction accuracy without trading off interpretability (Rai 2020).

*Methodological Procedure to Train, Interpret, and Validate a Mind Perception Classifier*

Using solicited consumer reviews, we train RoBERTa (Y. Liu et al. 2019). Once the pretrained model is fine-tuned with our training data, we evaluate the prediction accuracy on a hold-out test set, benchmarking it against traditional prediction methods (random forest, LIWC dictionary). We then leverage manifold methods to interpret ("white-box") the language model, assessing specific topics, tokens and terms signaling mind perception in consumer reviews. To increase ecological validity (van Heerde et al. 2021), we validate our model with real-world data, by applying it to a large set of unsolicited customer reviews scraped from a major online review site. Finally, we assess its transferability to a different category of smart objects. Figure 1 shows a schematic overview of our proposed methodological procedure. We will discuss each step in greater detail next.

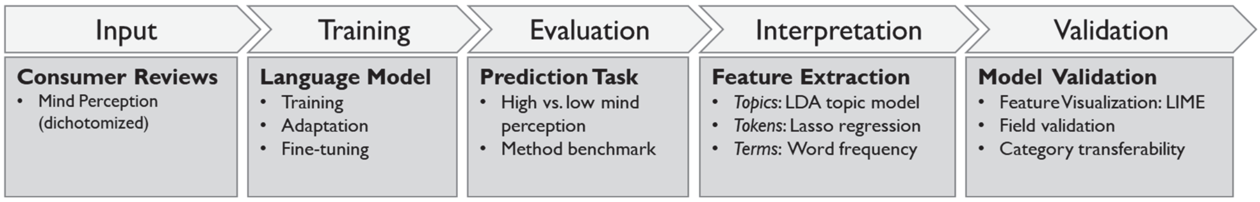
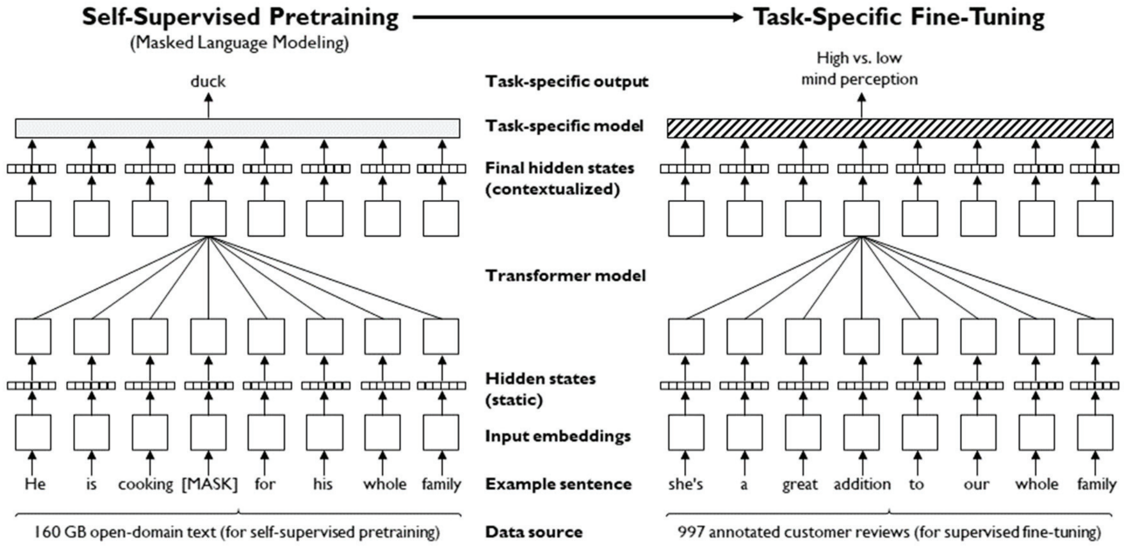



**Figure 5-1:** Methodological procedure to train, interpret, and validate a mind perception classifier

*Training & Evaluation.* In the first step, the language model is pretrained on massive open-domain text data, i.e., 160 GB of text, including entire English Wikipedia, for RoBERTa (Y. Liu et al. 2019)The language model is trained via self-supervised learning. In contrast to supervised learning, no human-annotated labels are required. Instead, words from the open-domain text are masked, and the language model is trained to predict these masked words correctly (so-called *masked language modeling*, Manning et al. 2020). In the second step, the architecture of the language model is adapted. Specifically, the task-specific head used for the language modeling task (i.e., last layer of the deep neural network) is replaced by a linear classifier. As shown in Figure 2, the linear classifier takes as input the pooled word embeddings (i.e., hidden states) of a document, which are returned by the language model, and projects them to the classification outcome, i.e., high mind perception vs. low mind perception. The parameters of this linear classifier are randomly initialized. Based on the training data the parameters of the linear classifier are estimated and the pre-trained parameters of the deep neural network are fine-tuned. Thereby, in encoding customer reviews, the pretrained language model can leverage the "knowledge" (e.g., word relations) from outside the training data, which it learned in the pre-training phase. While fine-tuning the language model is data efficient (797 observations in our case, 80% of the annotated data), it still requires computational resources that exceed the capacity of average desktop computers. Hence, we train the language model in the cloud on a specialized Tesla P4 graphics processing unit (GPU).




*Note*: Visualization adapted from (Wolf 2020)

**Figure 5-2:** Detailed overview of deep learning classification task

After fine-tuning the language model on the training data, we evaluate its prediction accuracy on the unseen hold-out test data (remaining 20% of the data, or in our case 200 observations). Specifically, we draw 50 bootstrapped samples from the test data to obtain an approximation for the variance of its classification performance.

Finally, in order to further evaluate model performance, we benchmark our model versus traditional classification methods. First, we train a random forest (Breiman 2001), which represents the popular class of decision tree ensembles and is well suited to automated text classification tasks (Hartmann et al. 2019). Second, given the popularity of lexicon-based methods in marketing and consumer research (e.g., Barasch and Berger 2014; Berger and Milkman 2012; Bradley and Meeds 2002; Kovács, Carroll, and Lehman 2014; Opoku et al. 2007) we also create a simple, dictionary-based model using LIWC as a feature extractor. Thus, this step allows us to assess the prediction accuracy of our model compared to other text classification approaches.

*Interpretation*. In order to address the transparency and interpretability limitations of a deep contextual language model, we propose using manifold methods to extract linguistic features associated with higher mind perception in customer reviews. Specifically, we aim to extract specific topics, terms, and tokens linked to mind perception in smart objects. To identify topics associated with high (as opposed to low) mind perception, we run a Latent Dirichlet Allocation (LDA, Blei, Ng, and Jordan 2003). LDA is a popular unsupervised machine learning method that is frequently used in marketing for exploratory text analysis (e.g., Büschken, Allenby, and Büschken 2016) and considered particularly useful for insight generation and interpretation (Berger et al. 2020). An LDA topic model uses a bag-of-word approach to identify words that commonly co-occur in a text, and thus define a broader underlying theme. By assessing the distribution of customer reviews classified as high (as opposed to low) mind perception across the LDA topics, this analysis provides insight into which topics are associated with higher mind perception.

Next, we use a logistic regression with L1 penalization (Lasso regression) to extract tokens (i.e., unigrams and bigrams) associated with high (as opposed to low) mind perception. The penalization in the loss function of the Lasso regression shrinks most of the regression coefficients to zero. The remaining terms represent the ones with the strongest correlation between the tokens and the degree of mind perception (high vs. low), thus allowing us to isolate specific tokens associated with higher mind perception.

While the first two analyses (LDA and Lasso regression) are exploratory, aimed at uncovering specific linguistic features (topics and tokens) associated with higher mind perception, the third step aims at confirming the incidence of specific terms in high compared to low mind perception reviews. Specifically, we take individual terms (individual, stemmed words) revealed by the topic modelling and Lasso repression and compare the frequency of their incidence in high (as opposed to low) mind perception reviews to further substantiate our insights and confirm specific terms associated with higher mind perception in customer reviews. Taken together, these three analyses provide transparency on which linguistic features are associated with high mind perception in customer reviews and thus provide an interpretable explanation for our model predictions.

*Validation*. Finally, our last step is aimed at validating our model, both, in terms of the specific linguistic features defining mind perception in smart objects, as well as generalizing our model with real-world field data, and assessing its transferability to a different category of smart objects. In order to validate key linguistic features of mind perception we leverage LIME (Local Interpretable Model-agnostic Explanations), a recently developed algorithm designed to explain classification predictions by approximating them locally with an interpretable model (Ribeiro, Singh, and Guestrin

2016). The objective of this approach is to increase trust in the classifier's predictions by visualizing features associated with the model predictions (Ribeiro et al. 2016; Rai 2020). While the classification function $f(x)$ that RoBERTa learns to distinguish low from high mind perception reviews is highly complex, LIME seeks to approximate this globally non-linear decision boundary with a linear model locally, i.e., in the vicinity of an instance of interest (see Figure A3 in the Web Appendix for an illustration). Instead of the original (non-interpretable) representation of the text in terms of contextual word embeddings, LIME represents the reviews using an interpretable bag-of-words approach. Hence, reviews are represented by the presence or absence of words. These serve as independent variables for the linear model with the confidence scores of a review being low compared to high mind perception as the dependent variable. Therefore, the weights assigned to words can be intuitively interpreted as the change in the predicted confidence that a review is low compared to high mind perception if a particular word is removed. The linear model is estimated on random samples of the data that are perturbed and weighted by their distance to the example to be interpreted. This renders it possible to evaluate and visualize the importance of individual components of a customer review (i.e., the words) in terms of the degree of expressed mind perception, and thus enables validating the linguistic features extracted in the previous step.

Finally, in order to assess the generalizability of our model, we take a two-step approach. First, we validate our model with real-world consumer reviews, scraped from a major online review site (Amazon.com). Leveraging our model prediction and LIME, we are able to assess whether the linguistic features associated with high mind perception generalize across unsolicited consumer reviews. Second, we validate our model predictions by applying it to a different category of smart objects in order to assess its generalizability and transferability. Again, we use LIME to interpret our model predictions and assess whether our extracted linguistic features are consistent with mind perception theory and generalize across categories.

Taken together, our proposed methodological procedure aims at providing a mind perception classifier that distinguishes itself based on three key characteristics. First, using RoBERTa, we aim to provide a classifier that has a higher *accuracy* in predicting mind perception from unstructured text compared to other, more established text classification methods. This is not an end in itself but necessary due to the difficulty of the task and subtle signals we need to extract from unstructured, unsolicited customer reviews. Second, we aim to provide a classifier with *high interpretability*, allowing the extraction of specific linguistic features associated with higher mind perception in smart objects. Finally, we aim to provide a classifier with *high external and ecological validity* (van Heerde et al. 2021) that generalizes across a broad set of smart objects. As such, our proposed approach aims to answer recent calls to push methodological boundaries bringing together multiple analytical methods on different data types to provide novel insight on an emerging technological phenomenon (Schmitt et al. 2021)

## Overview of Studies

In what follows we will present empirical work aimed at training and evaluating a mind perception classifier based on the procedure outlined above, and demonstrate how this novel approach can shed light on the influence of mind perception on consumer-smart object relationships and on the extent of task delegation. As outlined in Figure 3, we will present evidence from three studies to support our theorizing and demonstrate the advantages of our methodological procedure. Specifically, we show that our deep contextual learning model (compared to other text classification methods) does not only provide a higher accuracy level for the classification of mind perception in customer reviews, but that our proposed manifold methods enable the extraction of a rich set of linguistic features, which provide deeper insight into the underlying consumer-smart object relationships. Furthermore, we show that mind perception in consumer reviews is an important predictor of the extent of tasks delegated to smart objects, thus supporting the power of such a model to gain new insights into specific consumer behavior.

In Study 1 we train our language model based on solicited customer reviews from an extensive survey and evaluate the performance of our model on a hold-out test data, benchmarking it against two alternative classification models. Using manifold methods, we show how specific linguistic features can be extracted from such a model and which topics, terms, and tokens are associated with higher mind perception. Furthermore, we provide evidence for the importance of mind perception as a predictor of task delegation to smart objects (Study 1). In our field studies we apply our text classifier to real-world customer reviews scraped from an online review site in order to further demonstrate its validity and generalizability to unsolicited customer reviews (Study 2), as well as assess its transferability to an entirely different category of smart objects (Study 3a and 3b).

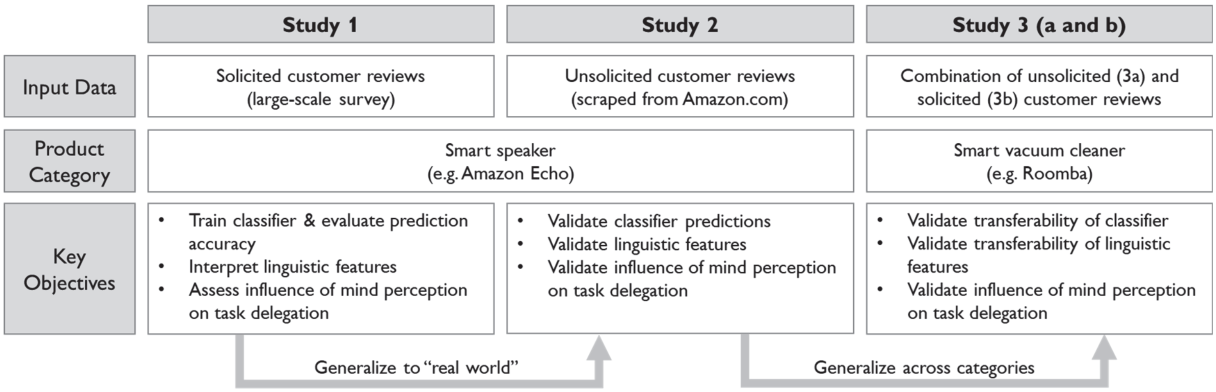



**Figure 5-3:** Overview of studies and key objectives

## Study 1: Training and Evaluation of Mind Perception Classifier

Study 1 was designed to train and evaluate a text classifier using state-of-the-art transfer learning to predict mind perception in a smart object from unstructured customer reviews. Furthermore, the study aims at extracting specific linguistic features associated with high mind perception to shed light on the underlying consumer-smart object relationships. Finally, Study 1 assesses whether the extent of mind perception predicts differences in smart-object usage behavior, i.e., the number of tasks consumers delegate to a smart object.

### *Sample and Procedure*

A total of 2509 active smart speaker users were recruited from an online consumer panel ($M_{age}$ = 35.3; 57.4% female). 14 participants were excluded due to missing data, leaving 2495 participants for the analysis. Participants specified the brand of the smart speaker they own, how long they have owned it, how often they interact with it on a daily basis, how many separate tasks they execute with it, and how many other devices are connected to it. Participants were then instructed to write a review about their experience using their smart speaker, imagining they were writing an online review about it (for example on Amazon.com). No other instructions were given and participants were free to write as long as they wished (although we asked them to write at least 50 words, in order to make sure we get sufficient text data for our classifier). Finally, in order to train the classifier on labeled data, we measure the extent of participants' mind perception of the smart speaker using 8 mind perception dimensions (all items adapted from Gray et al. 2007; sample items: "the smart speaker is able to convey thoughts and feelings", "the smart speaker is able to have experiences and be aware of things"; $\alpha$=.90, see Web Appendix A5).

*Mind Perception as a Predictor of Task Delegation*

*Textual and Usage Data*. Before training the text classifier, i.e., RoBERTa, we first assessed whether mind perception in smart objects predicts differences in the number of tasks delegated to it. We used participants' self-rated mind perception score as our primary predictor variable for task delegation. We also added variables related to participants' review text and current usage behavior to assess whether mind perception would relate to task delegation beyond what these variables could explain. Specifically, the usage data covered the brand consumers currently own (dummy coded as Amazon = 1, all other = 0) to account for any differences in the number of available skills across different brands, the duration of how long consumers own their device (in months) to account for differences in usage familiarity, the number of other smart devices the speaker is connected to and their self-reported daily interaction frequency with the device. The textual data included the number of words in the review to control for differences in review lengths linked to enumerating different task. Our dependent variable was the number of different tasks participants indicate they delegate to the device overall. Table 1 provides an overview of the descriptive statistics for these variables.

**Table 5-1:** Descriptive statistics (Study 1)

| Variables | Min | Max | Mean | SD | Freq |
|---|---|---|---|---|---|
| **Usage** | | | | | |
| Brand (1=Amazon) | | | | | .73 |
| Time owned (in months) | .50 | 72.00 | 19.32 | 12.16 | |
| Daily interaction frequency | 1.00 | 51.00 | 5.51 | 5.48 | |
| Number of connected devices | 1.00 | 12.00 | 2.49 | 2.09 | |
| Number of delegated tasks | 1.00 | 51.00 | 4.41 | 3.62 | |
| **Textual** | | | | | |
| Number of words in review text (in 100) | .07 | 3.81 | .90 | .35 | |
| **Self-assessment scales** | | | | | |
| Mind Perception (7-pt Likert-scale) | 1.00 | 7.00 | 2.84 | 1.30 | |

*Results*. The OLS regression analysis, using our independent variables (usage, textual, and self-assessed mind perception) as our predictor variables and the number of tasks delegated to the smart speaker (M = 4.41, SD = 3.62) as our dependent variable is significant ($F(6, 2486) = 148.3$, $p < .001$, $R^2 = .264$). In line with our hypothesis and displayed in Table 2, we find that greater mind perception is significantly associated with a higher number of tasks delegated to the smart speaker ($\beta = .17$, $t = 5.467$, $p < .001$), above and beyond what would be predicted by their daily interaction frequency ($\beta = .27$, $t = 21.615$, $p < .001$) and the number of connected devices ($\beta = .29$, $t = 8.891$, $p < .001$). Brand, time owned and the number of words in the review text all did not significantly predict the number of tasks delegated to the device. This result suggests that mind perception may capture a consumer perception of the device that goes beyond mere usage of the speaker.

**Table 5-2:** Mind perception as a predictor of task delegation

| | Dependent Variable: |
|---|---|
| | **Number of delegated tasks** |
| Mind Perception | .17*** |
| Time Owned | .01 |
| Daily Interaction Frequency | .27*** |
| Number of Connected Devices | .29*** |
| Review Word Count (in 100) | -.07 |
| Brand (1=Amazon) | -.13 |
| Constant | 1.72*** |
| Observations | 2493 |
| $R^2$ | .26 |

*Note:* *$p < .05$; **$p < .01$; ***$p < .001$

*Training the Classifier*

*Data and Method.* Next, we trained an automated text classifier to detect mind perception in the review text. This is a non-trivial task as we did not specifically instruct consumers to describe their relationship with their smart speaker in any particular way, but instead, simply asked them to write a review about their experience with the smart speaker. Our primary variable to train the mind perception classifier on labeled data was consumers' self-assessed mind perception rating (M = 2.84, SD = 1.30). We dichotomize this target variable, taking the top and bottom 20% of reviews as the positive and negative class respectively (see Pryzant, Chung, and Jurafsky (2017) for a similar approach). As outlined in the method section, we employ a state-of-the-art transfer learning classifier to leverage linguistic patterns and relationship outside the training data (see previous section on method, as well as Web Appendix A1 for further details). Specifically, we fine-tune RoBERTa, a pretrained deep contextual language model (Y. Liu et al. 2019). We formulate the mind perception prediction as a binary classification problem by dichotomizing the data, using the top and bottom 20% of observations of the mind perception distribution (Pryzant et al. 2017). Hence, the baseline accuracy is 50%. In total, 997 observations remain. We use 80% of the data for training the model and 20% as hold-out test data.

*Evaluating the Classifier Prediction Accuracy.* After fine-tuning the language model on the training data (N=797, 80%), we evaluate it on the unseen hold-out test data (N=200, 20%). Specifically, we draw 50 bootstrapped samples from the test data to obtain an approximation for the variance of the classification performance. This procedure yields an average hold-out accuracy of 75.4% (.46 SE).

*Convergent Validity.* Comparing the classifier prediction results and the self-reported mind perception rating (7-pt Likert scale) yields a significant correlation, confirming the internal consistency and validity of the model results ($r$ = .49, $t$(2491) = 28.095, $p$ < .001). Furthermore, estimating a generalized linear model with a logistic link function to assess whether consumers' average mind perception rating (claimed, 7-pt Likert scale) significantly predicts the classifier results (coded as 1 = high mind perception and 0 = low mind perception), also further confirmed the validity of the mind perception classifier to correctly identify mind perception from consumers' unsolicited review data ($z = 19.98$, $p < .001$).

*Method Benchmarking.* In order to demonstrate the superiority of RoBERTa compared to other, more traditional classification methods, we also employed two additional text classification models. First, we trained a random forest (Breiman 2001), which represents the popular class of decision tree ensembles and are well suited to automated text classification tasks (Hartmann et al. 2019). We specified a random forest using 500 decision trees, each trained on bootstrapped samples of the training data (so-called bagging, Breiman 1996) as well as taking a random subset of features at each decision tree split. The intention of these randomization techniques is to decorrelate individual decision trees, which improves overall ensemble performance. Unlike our language model, which represents words as word embeddings, our random forest takes as input a document-term-matrix using a bag-of-words approach that can reflect neither word order nor context-dependent meaning of words (see Web Appendix A1.2 for hyperparameters and details on pre-processing).

Second, given the popularity of lexicon-based methods in marketing and consumer research (e.g., Barasch and Berger 2014; Berger and Milkman 2012; Bradley and Meeds 2002; Kovács, Carroll, and Lehman 2014; Opoku et al. 2007) we also create a simple, LIWC-based model. Specifically, we use LIWC as a feature extractor (see (Netzer et al. 2019) for a similar approach). Feeding all observations from the training set and hold-out test set through LIWC, gives us a low-dimensional representation of each customer review in terms of the 93 LIWC lexicons, e.g., “social” or “family” (see Web Appendix 1.3 for a full list of all LIWC dimensions). For each sentence LIWC counts the total number of words and returns the share of words that are part of one of its lexicons, e.g., .17 for the “family” lexicon for the example sentence “Alexa is part of our family!”. We then estimate a logistic regression using the share of each lexicon as the independent and high versus low mind perception as the dependent variable, and evaluate it on the hold-out test set. Table 3 presents the method performance of RoBERTa compared to the two traditional benchmark methods, for which we also draw 50 bootstrapped samples of the test data to obtain estimates for the variance of their classification performance. This benchmarking exercise demonstrates that our RoBERTa classifier significantly outperforms the other two models by a margin of close to 10 percentage points.

**Table 5-3:** Comparison of RoBERTa to benchmark approaches

| Model | Prediction Accuracy | SD | SE |
|---|---|---|---|
| RoBERTa | 75.4 | 3.22 | .46 |
| Random forest | 65.5 | 3.44 | .49 |
| Dictionary (LIWC) based model | 63.8 | 3.73 | .53 |
| *Random chance (baseline accuracy)* | *50.0* | | |

*Note*: All values in %.

The results of the convergent validity analysis, as well as the comparison to a random forest and a LIWC-based model support both the robustness and also superiority of RoBERTa to detect subtle cues from text and provide significantly higher accuracy in predictions results.

*Interpretation: Extracting Linguistic Features of Mind Perception*

*Identifying Mind Perception Topics.* As outlined in our method section, we use topic modeling to try to identify and extract the underlying themes of high mind perception reviews. To this end, we ran an LDA topic model with the customer reviews as our text source. We use the same data subset, we used for our mind perception classifier training and evaluation (N=997). Leveraging a density-based method for adaptive LDA model selection (Cao et al. 2009) to determine the optimal number of topics, we found that a model with five topics performs best (see Web Appendix A2.1 for details). The five LDA topics with the most discriminative bigrams with highest relevance are displayed in Figure 4. As in factor analysis, LDA identifies the words that most often occur together and that therefore have the highest probability of being associated within a set of reviews. By observing the bigrams with the highest relevance for each topic, we can capture some local context of words and start making some inferences about the underlying theme within each topic (for a full list of the top 40 bigrams for each topic see Web Appendix A2.1). Specifically, the two first topics seem to significantly differentiate in their extent of the personification of the smart speaker. While in topic 1 the smart speaker is mainly referred to as “it” and “the device”, in topic 2 the smart speaker is referred to by name (“Alexa”) and “her”. Thus, in topic 1 the smart object is mainly treated as a simple device, while in topic 2, the smart object seems to be treated more as an entity. The second feature that distinguishes these topics can be found in comparing topics 3 and 4. While topic 3 suggests a more agentic relationship in which the consumer "uses" the smart object to complete a task, topic 4 suggests a more communal relationship in which the smart speaker "helps" the consumer to complete certain tasks. Finally, topic 5 seems to be primarily self-focus, including a majority of bigrams with "I".


*Note*: The terms are the most representative bigrams (λ=.5, see (Netzer et al. 2019)). The topic names are conceived by the authors based on the analysis of the most relevant bigrams as LDA does not automatically provide labels for the topics.

**Figure 5-4:** LDA topics and bigrams with highest relevance

*Topic Relevance for High vs. Low Mind Perception.* Next, we calculate the distribution of customer reviews across the five topics identified by the LDA topic model, combining it with the self-reported mind perception ratings to split the data into high and low in mind perception reviews. As illustrated in Figure 5, we find that topic 1 where the smart speaker is described as a device ("it", "the device") is significantly more prevalent in low mind perception as compared to high mind perception reviews ($P_{High} = 18\%$, $P_{Low} = 25\%$, $t = 6.110$, $p < .001$). We also find that topic 2 in which the smart speaker is referred to as an entity ("her", "alexa") occurs more frequently in high mind perception as opposed to low mind perception reviews ($P_{High} = 18\%$, $P_{Low} = 14\%$, $t = 4.834$, $p < .001$). Similarly, we find that topic 3 where the relationship with the smart speaker employs more agentic terms (e.g., "use it") is significantly more prevalent in low mind perception as compared to high mind perception reviews ($P_{High}$ = 23%, $P_{Low}$ = 26%, $t$ = 2.231, $p$ < .001). On the other side, we find that topic 4 in which the relationship with the smart speaker is referred to in more communal terms (e.g., "helps me") is more representative of high mind perception as opposed to low mind perception reviews ($P_{High}$ = 21%, $P_{Low}$ = 14%, $t$ = 6.582, $p$ < .001). Finally, topic 5 identified as a self-focused experience occurs to an insignificantly different degree in low and high mind perception reviews ($p$ = .112). This analysis provides initial insight that higher mind perception may be associated with linguistic features linked to the personification of the device (masculine and feminine pronouns as well as name of the entity) as well as to a more communal relationship (defined by the terms used to describe task execution).

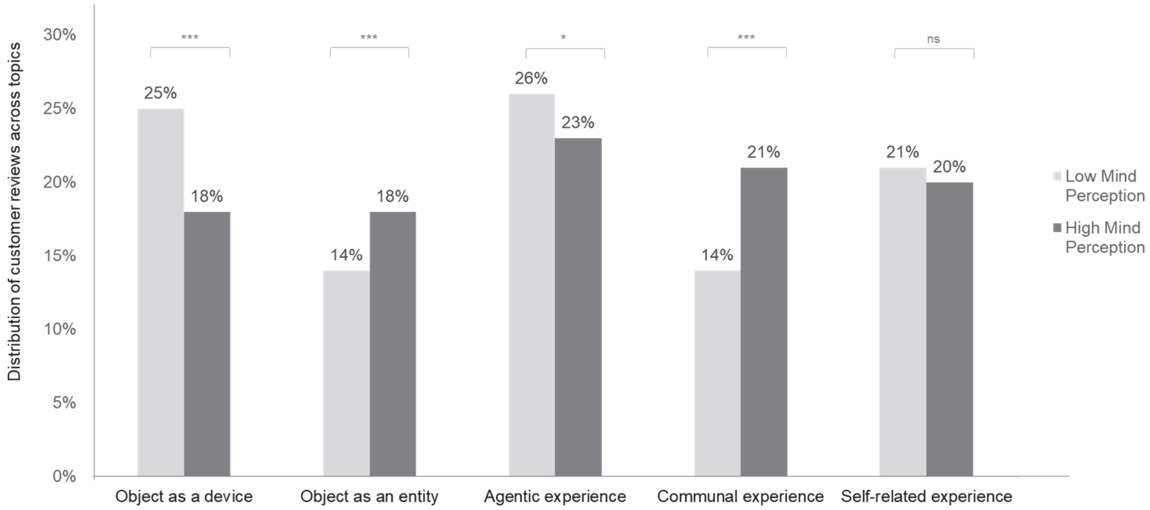
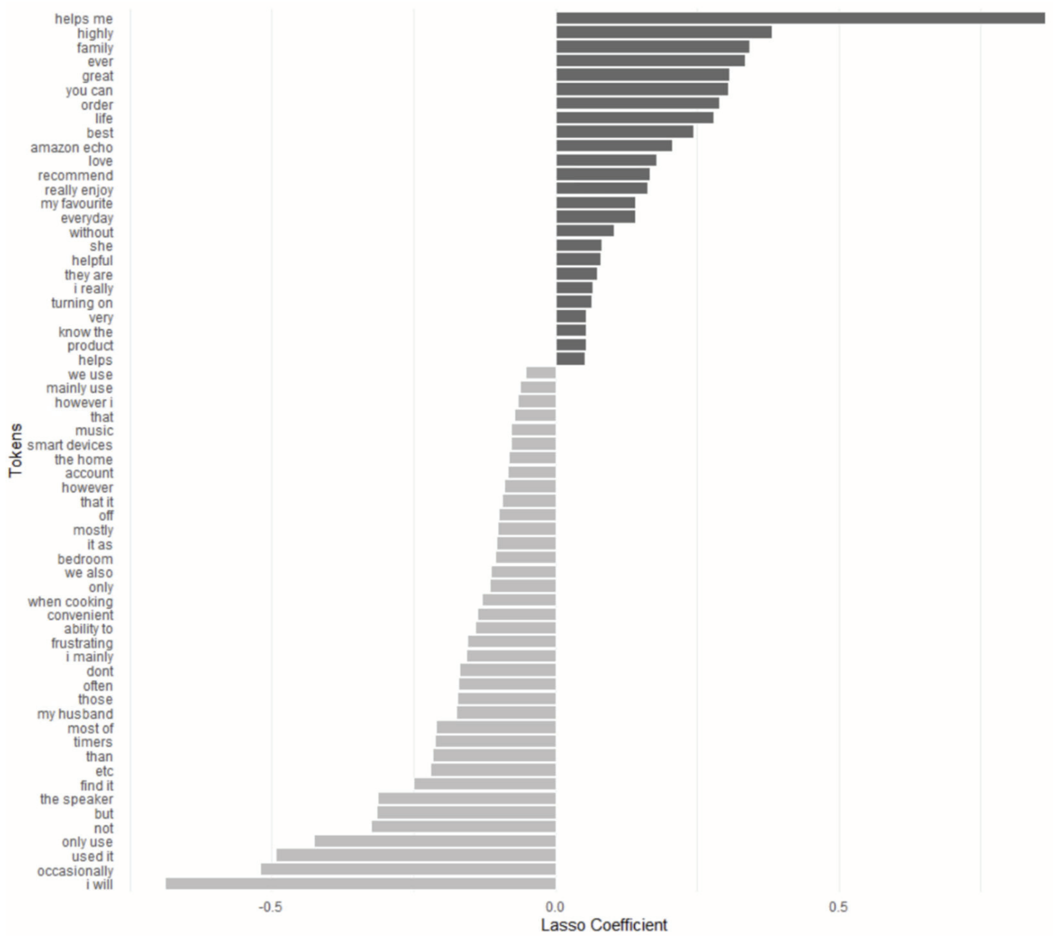



**Figure 5-5:** Distribution of high and low mind perception reviews across LDA topics

*Identifying Mind Perception Tokens.* To further confirm the insights gained through the LDA topic model, we ran a Lasso regression (a logistic regression with L1 penalization), using the dichotomized mind perception variable as the dependent variable (high = 1, low = 0) and a bag-of-words representation from the customer review data as the predictor variables (see Web Appendix A 2.2 for details on the preprocessing). The penalization in the loss function of the Lasso regression shrinks most of the regression coefficients to zero. The remaining ones are most representative and present the strongest correlation between the tokens (unigrams, bigrams) and the degree of mind perception (high vs. low). As can be seen in Figure 6, the results support our findings from the LDA topic model and show that those terms that are most representative of high mind perception describe the object as an entity ("she", "her") while those terms that are most representative of low mind perception describe the object as a device (“it”, “the speaker”, “smart device”). Furthermore, the Lasso regression also provides evidence for the difference in relationship we found in our LDA analysis. Specifically, those terms that are more representative of high mind perception are indicative of a more communal relationship (“helps me”, “family”), while those terms that are more representative of low mind perception are indicative of a more agentic relationship (“use it”).

Additionally, it appears as if high mind perception is partially driven by differences in sentiment, as evident by terms such as “great”, “best”, “love” in high mind perception reviews as opposed to “frustrating” in low mind perception reviews. This suggests that high mind perception may be associated with a more positively-valenced user experience. However, sentiment alone is not sufficient to predict differences in mind perception, as more subtle features are revealed through this analysis.




*Note:* Tokens with an absolute coefficient above .05 are included in this figure.

**Figure 5-6:** Predictive tokens for high mind perception using Lasso regression

*Identifying Mind Perception Terms*. Finally, comparing the incidence frequency of specific terms in high as compared to low mind perception reviews further corroborates our findings. Based on the previous analyses, we assessed term frequency for four terms, each selected to confirm the topics and tokens identified. Specifically, we assessed the incidence of “it” and “she” to confirm the extent of personification in high compared to low mind perception reviews. Furthermore, we assessed the incidence of “use” and “help” (both assessed as a stem, thus allowing for variations such as “useful”/ “helpful” or “used” / “helped” for example) to confirm the differences in smart object relationship. This analysis shows that the term “it” has a significantly higher incidence rate in low mind perception as compared to high mind perception reviews ($P_{HighMP} = 81\%$, $P_{LowMP} = 90\%$, $\chi^2 = 15.098$, $p < .001$), while the term “she” has a significantly higher incidence rate in high mind perception as compared to low mind perception reviews ($P_{HighMP} = 18\%$, $P_{LowMP} = 12\%$, $\chi = 7.8116$, $p < .01$), further supporting the difference in the extent of personification of the device. Furthermore, the term “use” has a significantly higher incidence rate in low mind perception as compared to high mind perception reviews ($P_{HighMP} = 72\%$, $P_{LowMP} = 80\%$, $\chi = 8.646$, $p < .01$), while the term “help” has a significantly higher incidence rate in high mind perception as compared to low mind perception reviews ($P_{HighMP} = 31\%$, $P_{LowMP} = 19\%$, $\chi = 18.996$, $p < .001$), further supporting our earlier findings that low mind perception reviews are characterized by a more agentic consumer-object relationship, while high mind perception reviews are characterized by a more communal consumer object relationship.

*Validating Linguistic Features: LIME*

To validate the linguistic features in our mind perception classifier, we leverage LIME (Ribeiro et al. 2016), a recently developed technique enabling a post-hoc explanation of a deep learning model by approximating the globally non-linear decision boundary of our RoBERTa classifier with a linear model locally. As detailed further in the method section, LIME is able to assign weights to individual words and to link them to the change in predicted confidence that a review is high compared to low mind perception. This renders it possible to evaluate and visualize the importance of individual words in terms of the degree of expressed mind perception, and thus enables validating the linguistic features extracted in the previous step and picked up by the language model.

To illustrate this step, we select two real customer reviews from our survey (introduced earlier as reviews from Customer A and Customer B, see page 1). which are comparable in terms of customer rating and length of review. We use our text classifier to classify the reviews as either high or low mind perception and apply LIME to generated salience maps to assess whether the model is picking up on the linguistic features signaling high mind perception consistent with our previous analytical step. As illustrated in Figure 7, the output from the LIME algorithm includes prediction probabilities for high as compared to low mind perception through its integration with our RoBERTa-based classifier. For the two exemplary reviews in Figure 7, there is a .95 probability that Customer A's review should be classified as representing high mind perception and a .94 probability that Customer B's review should be classified as low mind perception. The color coding represents the weights assigned to each word based on the change in the predicted probability that the review is low as opposed to high mind perception, and can thus be interpreted as the most representative features in those reviews. For example, if "use" were removed in the customer review of Customer B, then the probability of this review being classified as low mind perception would be $P_{LowMP}$ = .94 - .14 = .70. Similarly, if "help" were removed in the customer review of Customer A, then the probability of this review being classified as high mind perception would be $P_{HighMP}$ = .95 - .01 = .94. The low weights across the feature in review A further illustrate that even though "help" is recognized as a key feature by LIME, customer A's review contains many signals for high mind perception, thus even removing the term "help" does not significantly decrease the probability of this review being recognized as a high mind perception review.

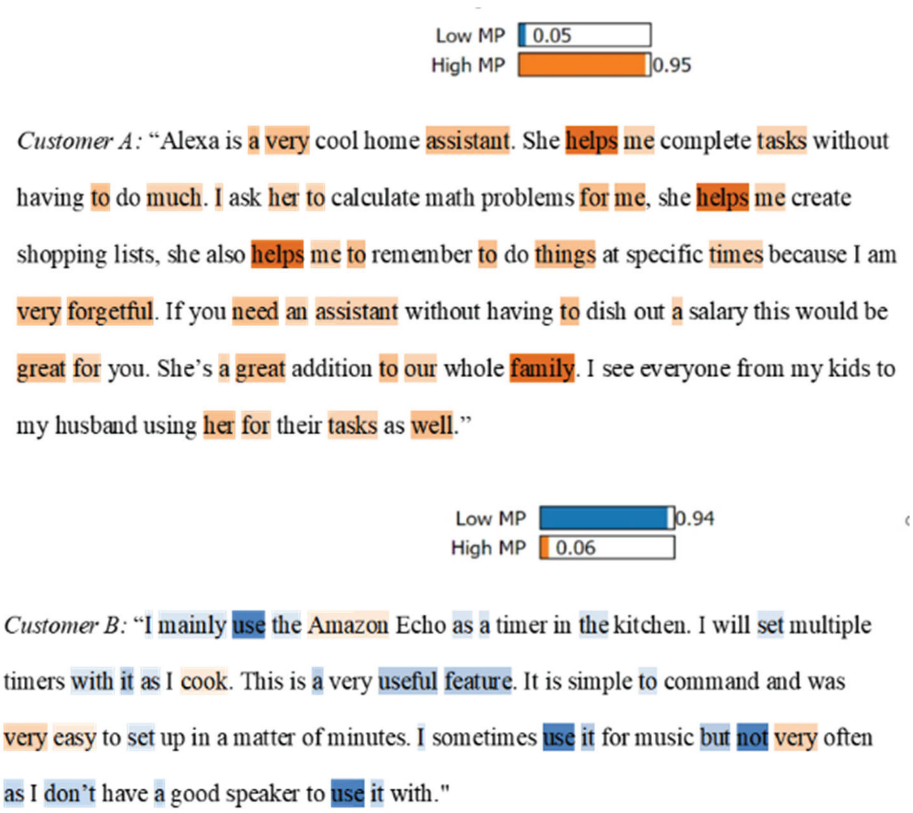



**Figure 5-7:** LIME salience maps (Study 1)

Supporting our previous findings on the key linguistic features, terms related to personification ("her") and a more communal relationship ("helps", "family", "ask") have the highest impact on predicting high mind perception, while terms that objectify the device ("it", "features") and reference a more agentic relationship ("use", "set") have the highest impact on predicting low mind perception for these reviews.

*Discussion*

The current study provides several new insights. First, we confirm that mind perception is an important predictor of the number of tasks active smart speaker users delegate to their assistant. Second, we train and validate a state-of-the-art deep contextual language model that can predict mind perception from unstructured customer reviews, and demonstrate its convergent validity with self-assessed mind perception, as well as its superiority to more traditional classification methods. Finally, we explore the topics, terms, and tokens associated with high and low mind perception, and uncover two key linguistic features distinguishing these customer reviews. Specifically, using a multi-method text interpretation approach that employs LDA topic modelling, lasso regression, and individual term frequency comparisons we find that high mind perception reviews are characterized by a stronger personification of the smart speaker, referring to it as an entity either by its name or by gendered pronouns, while low mind perception reviews refer to the smart speaker by gender-neutral pronouns and in a more objectified language. Second, we find evidence of differing consumer-object relationship roles between the high and low mind perception reviews. Specifically, the high mind perception reviews contain terms that are characteristic of a more communal relationship, while the low mind perception reviews contain terms that are characteristic of a more agentic relationship. Finally, integrating LIME with RoBERTa enables us to add an interpretable layer to our machine learning-based classifier providing a validation of the key linguistic features of higher mind perception. This new approach highlights the power of using deep contextual language models to uncover new consumer insights by designing an interpretable text classifier.

## Study 2: Field Validation

The objective of Study 2 is two-fold. First, the study is designed to evaluate the performance of our trained text classifier on unsolicited reviews, scraped from a major online review site (Amazon.com), and to assess whether the linguistic features for mind perception generalize among these real-world customer reviews. Furthermore, the aim of the study is to confirm whether mind perception predicts differences in consumers' willingness to delegate tasks to it.

### *Sample and Procedure*

For this study we scraped 20,000 Amazon Echo reviews from amazon.com. The reviews include the review text, author ID, review date, and review rating. We automatically classified whether the reviews mention the brand or product (e.g., "Amazon", "Echo") using a simple lexicon approach.

### *Prediction and model interpretation combining RoBERTa and LIME*

Using our text classifier (trained in Study 1) we classified all reviews as either high (coded as 1) or low (coded as 0) on mind perception. We then used LIME to create salience maps for 25 customer reviews to further assess whether the same linguistic features are salient for high as compared to low mind perception reviews. Specifically, we selected customer reviews that were comparable in length (e.g., number of words) and sentiment (4- or 5-star review ratings). We then used our text classifier to classify the reviews as high or low mind perception. Specifically, our RoBERTa-based mind perception classifier returns a probability score for each review between 0 and 1, which can be understood as the classifier's confidence in its prediction. LIME regresses these prediction probabilities on interpretable text features and highlights the most predictive words with warmer colors (orange vs. blue). Finally, we used the generated salience maps to assess whether the model was picking up features identifying the device as an entity (e.g., personal pronouns, names) and referring to a more communal relationship for high mind perception reviews consistent with our theorizing.

Figure 8 displays a selection of salience maps based on LIME for high and low mind perception reviews for illustration purposes. Supporting our previous findings on the key linguistic features, terms related to personification ("she" and "her") have a strong impact on predicting high mind perception. Additionally, LIME picks up on interesting, more advanced personification features in these reviews, such as a more active voice when describing the smart speaker's abilities and actions ("she reminds me", "she can answer", "she has a sense"). The algorithm also picks up on these terms in the low mind perception reviews, such as in the last review, where "lives" is highlighted as a feature that contributes to a higher mind perception, even though the rest of the review contains more low mind perception features. Our previous findings of a more communal relationship are also highlighted by the algorithm with the term "family" having a high impact on predicting higher mind perception, and going as far as "enhanced my life". On the other hand, terms that objectify the device either via a non-gendered pronoun ("it") as well as by referring to its functionalities ("alarm", "options") and terms referencing a more agentic relationship ("use", "program") have the highest impact on predicting low mind perception for these reviews. Thus, these salience maps further demonstrate how combining a RoBERTa-based classifier with a local interpretable algorithm such as LIME can provide rich insights into linguistic features linked to an underlying psychological mechanism such as mind perception.

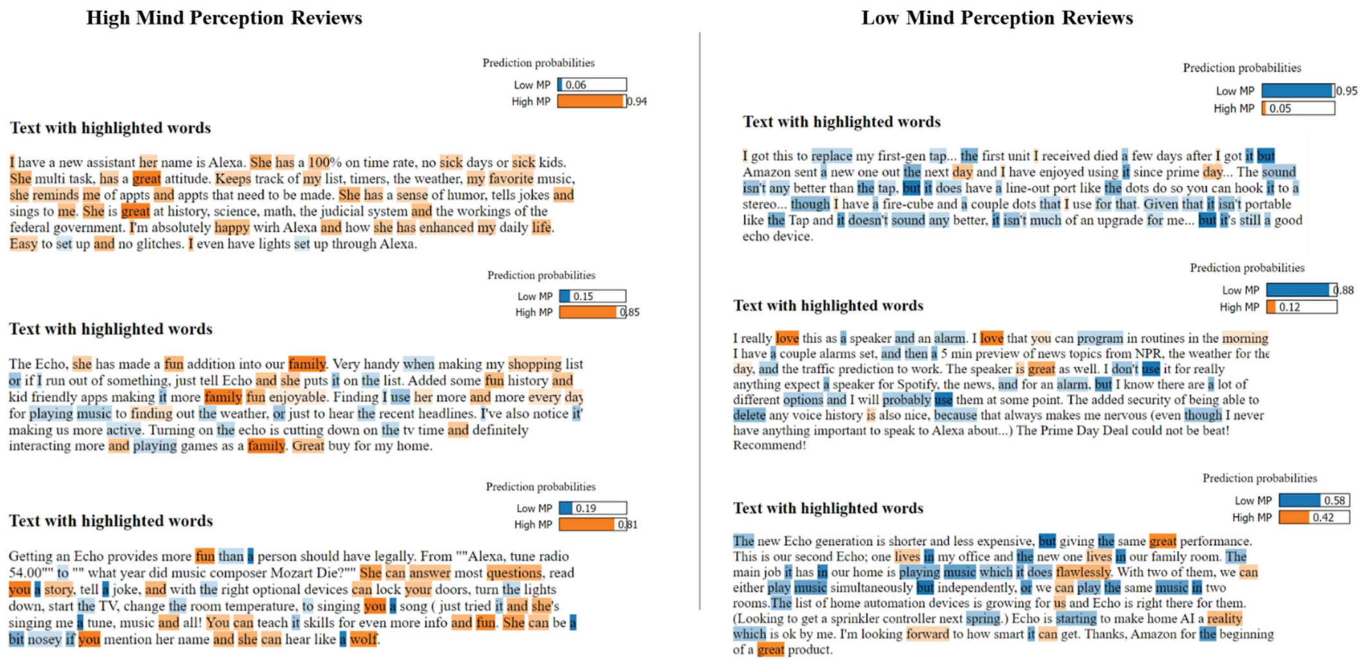



**Figure 5-8:** High and low mind perception reviews for a smart speaker with salience maps based on LIME

*Mind Perception as a Predictor of Task Delegation*

*Textual and Usage Data.* Next, in order to assess the extent of task delegation, two independent human coders who were blind to the hypothesis of our study manually coded the number of tasks mentioned in a subset of 500 (most recent) customer reviews. Specifically, the coders read through each of the reviews and counted the number of individual tasks consumers mentioned (e.g., "I use the Echo mainly to play music and control the lights" would be coded as two separate tasks). This number of individual tasks was our dependent variable. Mind perception (coded as 1 = high and 0 = low) as predicted by our classifier was our primary predictor variable. We also added variables related to participants' review text and usage behavior (as in Study 1) to assess whether mind perception would relate to task delegation beyond what these variables could explain. Specifically, the usage data covered the review age (in years) to account for differences in length of ownership and thus usage familiarity with the device. The textual data included the number of words in the review, to control for differences in review lengths linked to enumerating different task.

*Results.* An OLS regression model ($F(3, 496) = 38.61$, $p < .001$, $R^2 = .189$) with the predicted mind perception based on our classifier, as well as usage and textual data as predictor variables and the number of tasks mentioned by the consumer in the review as the dependent variable supports the validity of our text classifier. Specifically, as displayed in table 4, we find that mind perception significantly predicts the number of tasks consumers delegate to the smart speaker ($\beta$ = .94, $t$ = 6.007, $p$ < .001). Not surprisingly, the review length ($\beta$ = .74, $t$ = 9.459, $p$ < .001) is significantly associated with the number of delegated tasks, as we leveraged the number of tasks mentioned in the review text rather than self-claimed tasks as in Study 1. Finally, the review age is not significantly associated with the number of delegated tasks, suggesting that increasing familiarity with the device may not significantly influence consumers' willingness to delegate tasks to it.

**Table 5-4:** Mind perception predicts task delegation in customer reviews (Study 2)

| | Dependent Variable: |
|---|---|
| | Number of delegated tasks |
| Mind Perception (1=High) | .94*** |
| Review Word Count (in 100) | .74*** |
| Review Age (in Years) | .21 |
| Constant | .92** |
| Observations | 500 |
| $R^2$ | .26 |

*Note:* *$p$ < .05; **$p$ < .01; ***$p$ < .001

*Robustness check.* As an additional robustness check and given our findings that mind perception may be associated with a more positively-valenced consumer experience, we ran a regression model including sentiment as an additional predictor variable in our previous model. Specifically, we leveraged the AFINN lexicon (Nielsen 2011) to assign a sentiment score to each customer review. A significant OLS regression model ($F$(4, 495) = 34.04, $p$ < .001, $R^2$ = .216) confirms that high mind perception remains a significant predictor of task delegation ($\beta$ = .66, $t$ = 3.977, $p$ < .001), even though positive sentiment is significantly associated with higher task delegation ($\beta$ = .06, $t$ = 4.084, $p$ < .001).

*Discussion*

This field study confirms the importance of mind perception as a predictor of task delegation, and provides additional validation for our text classifier's ability to uncover subtle linguistic features in customer reviews that provide deeper insight into consumers' perception of and relationship with a smart speaker. It also provides additional evidence for the power of using deep contextual language models such as RoBERTa to uncover new consumer insights in unsolicited product reviews. To our best knowledge, this is the first paper to introduce LIME in conjunction with RoBERTa to the consumer behavior literature, adding an interpretable layer to a machine learning-based classifier that is commonly conceived as a black box (Rai 2020).

## Studies 3a and 3b: Transferability to Other Smart Objects

The objective of Study 3 is to gather initial evidence on whether the text classifier can transfer to a completely different domain of smart objects, a task that has a long tradition in NLP (e.g., Pan and Yang 2010), and which would further support the generalizability of our model and our approach. Specifically, Studies 1 and 2 focus on smart speakers, while in Study 3a and 3b we assess whether our text classifier can be transferred and generalized to another smart object: smart vacuum cleaners. Leveraging our procedural method for an interpretable text classifier applied in the previous studies, Study 3a uses the integration of LIME with our RoBERTa-based text classifier to categorize customer reviews and interpret the linguistic features of high (as opposed to low) mind perception reviews, while Study 3b aims at validating mind perception as a predictor of task delegation to the smart object.

### *3a: Sample and Procedure*

For this study we scraped 9,363 customer reviews on the smart vacuum cleaner iRobot Roomba from a major online review site (Amazon.com). The reviews included the review text, author ID, review date, and review rating. We used our text classifier trained during Study 1 to classify the reviews as high (coded as 1) or low (coded as 0) on mind perception. We then used LIME to create salience maps for 25 customer reviews in order to further assess whether the same linguistic features are salient for high as compared to low mind perception reviews. Specifically, we selected customer reviews that were comparable in length (e.g., number of words) and sentiment (4- or 5-star review ratings). Finally, we used the generated salience maps to assess whether the model was picking up features identifying the device as an entity (e.g., personal pronouns, names) and referring to a more communal relationship for high mind perception reviews.

*3a: Results*

Confirming our previous findings, the model appears to pick up on the same linguistic features as identified in Study 2 for smart speakers. Specifically, in the high mind perception reviews, the model highlights personal pronouns (e.g., “he”, “she”), pet names (e.g., “Ralph”, “Suzie”), as well as references to a more communal relationship with the smart vacuum cleaner (e.g., “member of the family”). To see a selection of salience maps from the LIME algorithm, please refer to the Web Appendix A4.

*Robustness Check.* As in our previous studies, the salience maps reveal that high mind perception is partially driven by differences in sentiment. While this can suggest that higher mind perception may be associated with a more positively-valanced consumer experience, it is clear from the salience maps that sentiment alone is not sufficient to predict differences in mind perception, as more subtle features are revealed through LIME.

*3a: Discussion*

Study 3a provides compelling first evidence that the text classifier trained on smart speaker reviews is transferable to a completely different category of smart objects (i.e., smart vacuum cleaners), and picks up on fundamental, generalizable linguistic features defining mind perception in smart objects. These results show the potential to identify mind perception among a completely different category of smart objects and further support and generalize our findings and theorizing on the key linguistic features defining higher mind perception in these devices. In order to confirm these results and assess the extent to which mind perception predicts task delegation in this new category of smart objects, we run a small follow-up survey among current owners of robot vacuum cleaners.

*3b: Sample and Procedure*

Mirroring the procedure of Study 1, a total of 100 active robot vacuum cleaner users were recruited from an online consumer panel ($M_{age}$ = 33.6; 49.5% female). 3 participants were excluded due to missing identification information, leaving 97 participants for the analysis. Participants specified the brand of robot vacuum cleaner they own, how long they have owned it for, how often they interact with it on a weekly basis, and how many other devices are connected to it. Participants were then instructed to write an online review about their robot vacuum cleaner (e.g., as they would on Amazon.com) and rate it on the same mind perception dimensions as used in Study 1 for the training of our classifier (all items adapted from Gray et al. 2007 $\alpha = .90$, see Web Appendix A5). As we are not able to assess task delegation for the autonomous vacuum cleaner based on current tasks mentioned in the online customer reviews as we did in Study 2 (given that it is designed to only execute one specific task, i.e., vacuuming), Study 3b asks consumers more directly about their willingness to delegate other tasks to the robot vacuum cleaner in the future. Specifically, we assessed delegation willingness by instructing them to imagine their robot vacuum cleaner could do other tasks and to indicate their likelihood to delegate each task to their robot vacuum cleaner (e.g., mow the lawn, rake leaves, collect parcels from the post office, and pick up food deliveries; 7-pt Likert scale, $\alpha = .91$).

*3b: Results*

*Evaluation of Classifier Prediction Accuracy*. We leverage our previously trained RoBERTa text classifier and again formulate the mind perception prediction as a binary classification problem by median-splitting the mind perception rating from the survey data. Hence, the baseline accuracy is 50%. This procedure yields an average accuracy of 61.03% for RoBERTa (.71% SE, based on 50 bootstrapped draws). Due to the limited sample size of only 97 open-text responses and the fact that our text classifier was not trained on reviews for robot vacuum cleaners specifically, our results can be considered a conservative but encouraging estimate as they show that the mind perception signals are transferable to an entirely different category of smart objects without any fine-tuning. In contrast, our baseline classifiers (random forest, LIWC; see Study 1) results in accuracy levels only marginally above random chance on the same hold-out test set, i.e., 50.10% (.58% SE) and 53.63% (.71% SE), respectively. These results once more emphasize the unique advantage of deep contextual language models such as RoBERTa, which can leverage generalizable language patterns from outside the training data through large-scale self-supervised pretraining (Manning et al. 2020).

*Mind Perception as a Predictor of Task Delegation.* In order to assess whether mind perception significantly predicts task delegation in a novel category of smart objects we run an OLS regression using the number of tasks participants are willing to delegate as our dependent variable (M = 3.17, SD = 1.99) and mind perception predicted by our classifier as our key predictor. In line with our previous models, we include usage data and textual controls in our model, to assess whether mind perception predicts task delegation beyond what these variables would predict. Specifically, we include the length of ownership in months, weekly interaction frequency with the device, and the number of other connected devices, as well as the review length as our textual control. Furthermore, given our findings in Study 3a that mind perception may be associated with a more positively-valenced consumer experience, also included sentiment score as an additional predictor variable in our previous model. Specifically, mirroring our approach in Study 2, we assess a sentiment score to each customer review using the AFINN lexicon (Nielsen 2011). Supporting our insight that higher mind perception is associated with higher task delegation, we find evidence that this result also holds in the case of robot vacuum cleaners. Specifically, as shown in Table 5, a significant OLS regression model ($F(6, 90) = 3.242$, $p < .01$, $R^2 = .178$) confirms that high mind perception is significantly associated with a higher likelihood to delegate additional tasks to the smart object ($\beta = 1.02$, $t = 2.103$, $p < .05$), while positive sentiment is not significantly associated with higher task delegation ($p = .38$).

**Table 5-5:** Mind perception predicts task delegation for smart vacuum cleaners

| | Dependent Variable: |
|---|---|
| | Number of delegated tasks |
| Mind Perception (1=High) | 1.02* |
| Time Owned | .04 |
| Daily Interaction Frequency | .04 |
| Number of Connected Devices | .41* |
| Review Word Count (in 100) | .58 |
| Sentiment Score | .04 |
| Constant | .57 |
| Observations | 97 |
| $R^2$ | .19 |

*Note:* *$p$ < .05; **$p$ < .01; ***$p$ < .001

*3b: Discussion*

Study 3b confirms the transferability of the text classifier to a different type of smart objects, i.e., smart vacuum cleaners. It also provides additional evidence for the relevance of mind perception to predict differences in consumers' willingness to delegate tasks to the smart object. Taken together, studies 3a and 3b provide converging evidence not only that our RoBERTa-based text classifier is transferable to other smart objects and can successfully identify linguistic markers of mind perception in unstructured, user-generated text, but also, that mind perception is an important predictor of consumer behavior, above and beyond consumer sentiment and usage data. This suggests that mind perception may be an important factor to consider when designing and evaluating consumer experiences with smart objects and its influence on downstream behavior.

## General Discussion

Automated text analysis has been called out as the next frontier in marketing to open up new avenues to gather consumer insights (Berger et al. 2020; Humphreys and Wang 2018; Schmitt 2019). In this study we propose a new methodological tool to conduct consumer research and a novel approach to theory development in marketing, leveraging a combination of diverse text analysis methods to classify customer reviews and extract linguistic features. Specifically, we show that blending state-of-the-art deep learning with traditional text mining approaches and local approximation algorithms for interpretable machine learning can open up new ways of conducting consumer research and gaining rich insights from unstructured text.

Across three studies we demonstrate how applying this procedure to online customer reviews can shed light on consumer perceptions and behaviors. Specifically, we show that higher mind perception fundamentally shapes the perception of smart objects, affecting their perceived relationship role, and ultimately, influencing consumers' willingness to delegate tasks to them. In Study 1 we provide evidence for the superiority of our deep learning classifier as compared to traditional automated text classification methods and are able to extract distinct linguistic features of high mind perception in unsolicited customer reviews. Specifically, the higher use of personal pronouns and communal relationship terms in these reviews appears to cast those smart objects in a role of an intelligent, empathetic entity rather than merely a lifeless device. Our two field studies further confirm the validity and transferability of our text classifier to detect subtle linguistic features in customer reviews that provide evidence of whether consumers attribute a higher mind to the smart object and its effect on downstream task delegation.

*Theoretical Contributions*

The current work makes three important theoretical contributions. First, this research is the first to propose a state-of-the-art text classifier to unobtrusively detect customers' perception of a mind in a non-human agent from unstructured review data, highlighting the power of deep contextual language models to detect more nuanced variations in text as compared to other classification methods or mere evaluations of consumer sentiment. This novel methodology provides an exciting opportunity to expand the experimental toolbox used in consumer research, and extends recent marketing literature on advancing automated text analysis to uncover behavioral insight (Berger et al. 2020; Humphreys and Wang 2018). Our research demonstrates the superiority of deep contextual language models to detect subtle linguistic features in textual data, and we use a series of analytical methods to unveil the distinct linguistic features of mind perception in the review text. This multi-method approach shows that even though deep contextual language models are often thought of as black boxes (Ribeiro et al. 2017, Rai 2020), by leveraging a series of analytical tools it is possible to shed light on the inner workings of such models, providing new opportunities to convert such unstructured customer reviews into valuable, actionable consumer insights.

Second, our insights contribute to and extend research on anthropomorphic agents in marketing, investigating a deeper level of anthropomorphism that goes beyond mere anthropomorphic appearance to perceiving a mind in non-human agents (Schmitt 2019; Yang, Aggarwal, and McGill 2020). This attribution of mental capacities and the ability of smart objects to dynamically interact with consumers and independently execute tasks on their behalf makes smart objects active partners in the consumption process. Thus, we contribute to emerging research on consumer-smart object relationships (Hoffman and Novak 2018; Novak and Hoffman 2018), providing evidence that perceiving a mind in an object can influence the perceived relationship role of the object and increase the extent to which consumers are willing to enable the object to execute tasks on their behalf, fundamentally shaping consumer experiences in the IoT.

Third, our findings contribute to the emerging field of technology-augmented choice (Melumad et al. 2020) demonstrating that smart objects become important modifiers of consumers' choice to the extent that consumers are willing to attribute responsibility to them. Our work demonstrates that an additional layer that may be considered in Melumad et al.'s (2020) framework for technology-augmented choice may lie in the subjective perception of the device itself. Thus, rather than just being a function of device type, the extent of mind perception in a device may be a further modifier to consider in technology-augmented decision-making, especially given the increasing intelligence, autonomy, and even empathy of such devices.

Given the influence of mind perception on consumer experience and choice, we ran a small pilot study in order to assess whether mind perception can be exogenously induced in a smart object through an interaction sequence. In this study, participants interacted with a smart speaker from a fictional brand called Yobu. Participants (N = 219, $M_{age}$ = 38.1; 45.1% female) and were randomly assigned to either a high mind perception condition (aimed at inducing high mind perception in Yobu) or a control condition. Building on the finding that non-human entities are low on subjective experience (Gray et al. 2007) the interaction aimed at inducing higher mind perception included questions about the smart speaker's subjective experience (e.g., "How are you?", "How was your day yesterday?"), perceptions (e.g., "How many people live in this house?") and preferences (e.g., "What is your favorite color?", "What makes you happy?"). In the control condition, participants asked the smart speaker factual questions, unrelated to the speaker's experience, such as "At what time is sunrise?", "How many people live in New York City?". In both conditions, consumers asked the smart speaker a total of ten questions as described above, and the interactions did not significantly differ in terms of length ($M_{HighMP}$ = 1026, $M_{Control}$ = 1001, $t(208) = 0.31$, $p = .8$). Participants then rated their perception of the smart speaker across several traits aimed at evaluating its perceived relationship role (e.g., sample items: "Yobu appears…" "caring", "helpful", "trustworthy", "understanding", $\alpha = .89$; Abele et al. 2008). Next, participants were provided with a range of tasks they could delegate, for which they indicated whether they would like to execute them independently (coded as 0), or whether they would ask the smart speaker to execute them on their behalf (coded as 1). For a full list of the tasks asked, please refer to Web Appendix A7. A one-way ANOVA revealed that participants form a more communal relationship with the speaker in the high mind perception condition ($M_{HighMP}$ = 5.13, $M_{Control}$ = 4.15, $F(1, 218) = 26.18$, $p < .001$)

and are more likely to delegate tasks to it ($t$(209=1.84, $p$ = .067). Furthermore a simple mediation model (5,000 bootstrap samples) reveals that high mind perception leads to a significant increase in the perception of a communal relationship with the speaker ($\beta_{Condition}$ = .97, t = 5.12, $p$ < .001), which in turn increases willingness to delegate tasks to it ($\beta_{Communal}$ = .05, t = 3.55, $p$ < .001), switching off the effect of the experimental manipulation ($p$ = .77) and leading to a significant indirect effect of mind perception on task delegation with a confidence interval excluding zero ($\beta$=.05, 95% CI [.02; .09]. Although these findings can only provide initial, preliminary evidence, they suggest that mind perception can be induced in smart objects by designing specific interaction sequences and demonstrate that they significantly affect not only how the speaker is perceived, but also influence its extent of usage more profoundly.

*Practical Contributions*

To the best of our knowledge, the current work is the first to provide an accessible tool to unobtrusively detect and interpret mind perception in customer reviews. The current findings demonstrate that the attribution of mental capacities to smart objects may not only shape consumers' perception of and engagement with the device, but may have far-reaching consequences on consumer experiences in the IoT. Indeed, in our Study 3a we also assessed whether mind perception can be used as a predictor of customer satisfaction with the device. A linear regression controlling for the length of reviews, product and brand mentions, review age, and sentiment score shows a significant impact of mind perception detected in the review text on customer ratings ($\beta$ = 1.88, $t$ = 3.94, $p$ < .001). This follow-up analysis not only demonstrates the transferability of our text classifier to other smart objects, but provides a novel tool for practitioners to unobtrusively assess the interaction experiences, potentially enabling them to build technology that is more engaging and relatable from a consumer perspective. This may have important implications across a broad set of industries and applications such as medical AI or driverless vehicles.

*Future Research*

One limitation of the current work is that we are using customer reviews due to the prevalence and accessibility of this textual data. However, a potentially interesting avenue for future research would be to use actual interaction data (i.e., conversational data) from direct consumer-smart object interactions. Given this voice data can be easily transcribed to written text, our proposed methodology could be applied to such kind of unstructured data. Furthermore, our methodological approach contributes to the recent stream of research on interpretable machine learning and how it can be used to understand novel phenomena and advance theory. It may extend beyond voice data to other sources of unstructured, user-generated data such as images or video for example. As these sources of data are increasingly used in generating consumer insights (Liu, Dzyabura, and Mizik 2020) our proposed methodology may provide novel opportunities to extract consumer insights from unstructured data without using human judges. For example, employing LIME or related methods such as gradient-weighted class activation mappings (Hartmann et al. 2021) can open-up exciting new opportunities to gain novel insights from consumer-generated data.

Finally, given the nature of smart speakers, we used task delegation as our primary behavioral outcome variable. However, it is highly likely that there are other behavioral consequences of a closer, more personal relationship with smart objects, such as in the context of human replacement (Granulo et al. 2019; Waytz and Norton 2014), personal cognitive enhancements (Castelo, Schmitt, and Sarvary 2019) or privacy concerns (Lau, Zimmerman, and Schaub 2018) for example. Thus, we believe that our mind perception classifier and our proposed methodological approach provide a vast and rich opportunities for future research on the underlying psychology of consumer-technology relationships and its consequences on consumer behavior.

*Conclusions*

The current work proposes a novel approach leveraging the latest technological innovations in natural language processing to uncover and interpret nuanced linguistic signals in unstructured text, opening up new avenues to conduct research and develop theory in marketing. The current paper focuses on mind perception as an application to demonstrate how the words consumers write provides an unexplored window into the depths of these subjective interaction experiences, allowing both researchers and practitioners to uncover novel insights into emerging consumer-smart object relationships.

However, mind perception is just one of many potential applications and our approach may be extended to unstructured, user-generated data beyond textual reviews. We demonstrate that using cutting-edge deep learning models and blending them with more interpretable algorithms provide innovative research tools and enable new theory development and richer insights on established psychological processes.

## References


Abele, Andrea E., Mirjam Uchronski, Caterina Suitner, and Bogdan Wojciszke (2008), "Towards an Operationalization of the Fundamental Dimensions of Agency and Communion: Trait Content Ratings in Five Countries Considering Valence and Frequency of Word Occurrence," *European Journal of Social Psychology*, 38, 1202–17.

Abell, F., F. Happé, and U. Frith (2000), "Do Triangles Play Tricks? Attribution of Mental States to Animated Shapes in Normal and Abnormal Development," *Cognitive Development*, 15(1), 1–16.

Aggarwal, Praveen and Tridib Mazumdar (2008), "Decision Delegation: A Conceptualization and Empirical Investigation," *Psychology & Marketing*, 25(1), 71–93, http://eprints.lancs.ac.uk/23431/.

Airenti, Gabriella (2018), "The Development of Anthropomorphism in Interaction: Intersubjectivity, Imagination, and Theory of Mind," *Frontiers in Psychology*, 9, 1–13.

Astington, Janet Wilde and Jodie A Baird (2005), *Why Language Matters for Theory of Mind*, New York, NY, US: Oxford University Press.

Barasch, Alixandra and Jonah Berger (2014), "Broadcasting and Narrowcasting: How Audience Size Affects What People Share," *Journal of Marketing Research*, 51(3), 286–99.

Batson, C. Daniel, Cynthia L. Turk, Laura L. Shaw, and Tricia R. Klein (1995), "Information Function of Empathic Emotion: Learning That We Value the Other's Welfare," *Journal of Personality and Social Psychology*, 68(2), 300–313.

Berger, Jonah, Ashlee Humphreys, Stephan Ludwig, Wendy W. Moe, Oded Netzer, and David A. Schweidel (2020), "Uniting the Tribes: Using Text for Marketing Insight," *Journal of Marketing*, 84(1), 1–25.

Berger, Jonah and Katherine L. Milkman (2012), "What Makes Online Content Viral?," *Journal of Marketing Research*, 49(2), 192–205.

Blei, David M, Andrew Y Ng, and Michael I Jordan (2003), "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, 3, 993–1022.

Bovet, Alexandre, Flaviano Morone, and Hernán A. Makse (2018), "Validation of Twitter Opinion Trends with National Polling Aggregates: Hillary Clinton vs Donald Trump," *Scientific Reports*, 8(1), 1–16, https://doi.org/10.1038/s41598-018-26951-y.

Bradley, Samuel D. and Robert Meeds (2002), "Surface-Structure Transformations and Advertising Slogans: The Case for Moderate Syntactic Complexity," *Psychology and Marketing*, 19(7–8), 595–619.

Breiman, Leo (1996), "Bagging Predictors," *Machine Learning*, 24, 123–40.

——— (2001), "Random Forests," *Machine Learning*, 45(1), 5–32, https://doi.org/10.1023/A:1010933404324.

Bretherton, Inge and Marjorie Beeghly (1982), "Talking about Internal States: The Acquisition of an Explicit Theory of Mind," *Developmental Psychology*, 18(6), 906–21.

Bretherton, Inge, Sandra McNew, and Marjorie Beeghly-Smith (1981), "Early Person Knowledge as Expressed in Gestural and Verbal Communication: When Do Infants Acquire a 'Theory of Mind'?," in *Infant Social Cognition*, ed. M. E. Lamb and L. R. Sherrod, Hillsdale, NJ: Erlbaum, 333–73.

Büschken, Joachim, Greg M Allenby, and Joachim Büschken (2016), "Sentence-Based Text

Analysis for Customer Reviews Sentence-Based Text Analysis for Customer Reviews," (December 2020).

Cao, Juan, Tian Xia, Jintao Li, Yongdong Zhang, and Sheng Tang (2009), "A Density-Based Method for Adaptive LDA Model Selection," *Neurocomputing*, 72(7–9), 1775–81.

Castelo, Noah, Bernd Schmitt, and Miklos Sarvary (2019), "Human or Robot? Consumer Responses to Radical Cognitive Enhancement Products," *Journal of the Association for Consumer Research*, 4(3), 217–30.

Delbaere, Marjorie, Edward F. McQuarrie, and Barbara J. Phillips (2011), "Personification in Advertising," *Journal of Advertising*, 40(1), 121–30.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova (2019), "BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding," http://arxiv.org/abs/1810.04805.

Dunn, Judy, Jane Brown, Cheryl Slomkowski, Caroline Tesla, and Lise Youngblade (1991), "Young Children's Understanding of Other People's Feelings and Beliefs : Individual Differences and Their Antecedents," *Child Development*, 62, 1352–66.

Epley, Nicholas, Adam Waytz, and John T. Cacioppo (2007), "On Seeing Human: A Three-Factor Theory of Anthropomorphism," *Psychological Review*, 114(4), 864–86.

Fournier, Susan (1998), "Consumers and Their Brands: Developing Relationship Theory in Consumer Research," *Journal of Consumer Research*, 24(4), 343–53, https://academic.oup.com/jcr/article-lookup/doi/10.1086/209515.

Gächter, Simon, Benedikt Herrmann, and Christian Thöni (2004), "Trust, Voluntary Cooperation, and Socio-Economic Background: Survey and Experimental Evidence," *Journal of Economic Behavior and Organization*, 55(4 SPEC.ISS.), 505–31.

Goldstein, Thalia R and Ellen Winner (2012), "Enhancing Empathy and Theory of Mind," *Journal of Cogntition and Development*, 13(1), 19–37.

Granulo, Armin, Christoph Fuchs, and Stefano Puntoni (2019), "Psychological Reactions to Human versus Robotic Job Replacement," *Nature Human Behaviour*, 3(10), 1062–69, http://dx.doi.org/10.1038/s41562-019-0670-y.

Gray, Heather M., Kurt Gray, and Daniel M. Wegner (2007), "Dimensions of Mind Perception," *Science*, 315(5812), 619.

Hartmann, Jochen, Mark Heitmann, Christina Schamp, and Oded Netzer (2020), "The Power of Brand Selfies in Consumer-Generated Brand Images," *SSRN*, Working Paper, https://dx.doi.org/10.2139/ssrn.3354415.

Hartmann, Jochen, Juliana Huppertz, Christina Schamp, and Mark Heitmann (2019), "Comparing Automated Text Classification Methods," *International Journal of Research in Marketing*, 36(1), 20–38, https://doi.org/10.1016/j.ijresmar.2018.09.009.

Haslam, Nick, Stephen Loughnan, Yoshihisa Kashima, and Paul Bain (2008), "Attributing and Denying Humanness to Others," *European Review of Social Psychology*, 19(1), 55–85.

van Heerde, Harald J., Christine Moorman, C. Page Moreau, and Robert W. Palmatier (2021), "Reality Check: Infusing Ecological Value into Academic Marketing Research," *Journal of Marketing*, 85(2), 1–13.

Heitmann, Mark, Christian Siebert, Jochen Hartmann, and Christina Schamp (2020), "More than a Feeling: Benchmarks for Sentiment Analysis Accuracy," *SSRN*, Working Paper, https://dx.doi.org/10.2139/ssrn.3489963.

Hills, Thomas T., Eugenio Proto, Daniel Sgroi, and Chanuki Illushka Seresinhe (2019), "Historical Analysis of National Subjective Wellbeing Using Millions of Digitized Books," *Nature Human Behaviour*, 3, 271–1275, http://dx.doi.org/10.1038/s41562-019-0750-z.

Hinds, Pamela J., Teresa L. Roberts, and Hank Jones (2004), "Whose Job Is It Anyway? A Study of Human-Robot Interaction in a Collaborative Task," *Human-Computer Interaction*, 19(1–2), 151–81.

Hirschberg, Julia and Christopher D Manning (2015), "Advances in Natural Language Processing," *Science*, 349(6245), 261–66.

Hoffman, Donna L. and Thomas P. Novak (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.

Howard, Jeremy and Sebastian Ruder (2018), "Universal Language Model Fine-Tuning for Text Classification," *arXiv preprint arXiv:1801.06146v5*.

Humphreys, Ashlee and Rebecca Jen Hui Wang (2018), "Automated Text Analysis for Consumer Research," *Journal of Consumer Research*, 44(6), 1274–1306, https://doi.org/10.1093/jcr/ucx104.

Komiak, Sherrie Y X and Izak Benbasat (2006), "The Effects of Personalization Trust and Adoption of Recommendation," *MIS Quarterly: Management Information Systems*, 30(4), 941–60.

Komiak, SYX, W Wang, and I Benbasat (2004), "Trust Building in Virtual Salespersons versus in Human Salespersons: Similarities and Differences," *e-Service Journal*, 3(3), 49–64.

Kovács, Balázs, Glenn R. Carroll, and David W. Lehman (2014), "Authenticity and Consumer Value Ratings: Empirical Tests from the Restaurant Domain," *Organization Science*, 25(2), 458–78.

Kryvasheyeu, Yury, Haohui Chen, Nick Obradovich, Esteban Moro, Pascal Van Hentenryck, James Fowler, and Manuel Cebrian (2016), "Rapid Assessment of Disaster Damage Using Social Media Activity," *Science Advances*, 2(3), https://doi.org/10.1126/sciadv.1500779.

Landwehr, Jan R, Ann L Mcgill, and Andreas Herrmann (2011), "It' s Got the Look : The Effect of Friendly and Aggressive "Facial" Expressions on Product Liking and Sales," *Journal of Marketing*, 75, 132–46.

Lau, Josephine, Benjamin Zimmerman, and Florian Schaub (2018), "Alexa, Are You Listening? Privacy Perceptions, Concerns and Privacy-Seeking Behaviors with Smart Speakers," *Proceedings of the ACM on Human-Computer Interaction*, 2, https://doi.org/10.1145/3274371.

Liu, Liu, Daria Dzyabura, and Natalie Mizik (2020), "Visual Listening in: Extracting Brand Image Portrayed on Social Media," *Marketing Science*, 39(4), 669–86.

Liu, Xiao, Dokyun Lee, and Kannan Srinivasan (2019), "Large-Scale Cross-Category Analysis of Consumer Review Content on Sales Conversion Leveraging Deep Learning," *Journal of Marketing Research*, 56(6), 918–43.

Liu, Yinhan, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov (2019), "RoBERTa: A Robustly Optimized BERT Pretraining Approach," *ArXiv preprint: arXiv:1907.11692*.

Ludwig, Stephan, Tom van Laer, Ko de Ruyter, and Mike Friedman (2016), "Untangling a Web of Lies: Exploring Automated Detection of Deception in Computer-Mediated

Communication," *Journal of Management Information Systems*, 33(2), 511–41, https://doi.org/10.1080/07421222.2016.1205927.

Manning, Christopher D., Kevin Clark, John Hewitt, Urvashi Khandelwal, and Omer Levy (2020), "Emergent Linguistic Structure in Artificial Neural Networks Trained by Self-Supervision," *Proceedings of the National Academy of Sciences*, 201907367.

Mead, George H (1934), *Mind, Self, and Society*, ed. Charles W. Morris, Chicago and London: University of Chicago Press.

Melumad, Shiri, Rhonda Hadi, Christian Hildebrand, and Adrian F. Ward (2020), "Technology-Augmented Choice: How Digital Innovations Are Transforming Consumer Decision Processes," *Customer Needs and Solutions*.

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean (2013), "Distributed Representations of Words and Phrases and Their Compositionality," *Journal of Machine Learning Research*, 3(4–5), 951–91, https://doi.org/10.1162/jmlr.2003.3.4-5.951.

Moon, Sangkil and Wagner A. Kamakura (2017), "A Picture Is Worth a Thousand Words: Translating Product Reviews into a Product Positioning Map," *International Journal of Research in Marketing*, 34(1), 265–85, http://dx.doi.org/10.1016/j.ijresmar.2016.05.007.

Netzer, Oded, Ronen Feldman, Jacob Goldenberg, and Moshe Fresko (2012), "Mine Your Own Business: Market-Structure Surveillance Through Text Mining," *Marketing Science*, 31(3), 521–43, https://doi.org/10.1287/mksc.1120.0713.

Netzer, Oded, Alain Lemaire, and Michal Herzenstein (2019), "When Words Sweat: Identifying Signals for Loan Default in the Text of Loan Applications," *Journal of Marketing Research*, 56(6), 960–80, https://doi.org/10.2139/ssrn.2865327.

Nielsen, Finn Årup (2011), "A New ANEW: Evaluation of a Word List for Sentiment Analysis in Microblogs," *CEUR Workshop Proceedings*, 718, 93–98.

Novak, Thomas and Donna L. Hoffman (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*, 216–37.

Novak, Thomas P. and Donna L. Hoffman (2019), "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), 216–37, https://doi.org/10.1007/s11747-018-0608-3.

Opoku, Robert Ankomah, Russell Abratt, Mike Bendixen, and Leyland Pitt (2007), "Communicating Brand Personality: Are the Web Sites Doing the Talking for Food SMEs?," *Qualitative Market Research*, 10(4), 362–74.

Ordenes, Francisco Villarroel, Dhruv Grewal, Stephan Ludwig, Ko De Ruyter, Dominik Mahr, and Martin Wetzels (2019), "Cutting through Content Clutter: How Speech and Image Acts Drive Consumer Sharing of Social Media Brand Messages," *Journal of Consumer Research*, 45(5), 988–1012.

Ordenes, Francisco Villarroel, Stephan Ludwig, Ko De Ruyter, Dhruv Grewal, and Martin Wetzels (2017), "Unveiling What Is Written in the Stars: Analyzing Explicit, Implicit, and Discourse Patterns of Sentiment in Social Media," *Journal of Consumer Research*, 43(6), 875–94.

Packard, Grant and Jonah Berger (2019), "How Concrete Language Shapes Customer Satisfaction," in *Advances in Consumer Research*, ed. Rajesh Bagchi, Lauren Block, and

Leonard Lee, Duluth, MN: Association for Consumer Research, 29–34.

Pan, Sinno Jialin and Qiang Yang (2010), “A Survey on Transfer Learning,” *IEEE Transactions on Knowledge and Data Engineering*, 22(10), 1345–59.

Parker, Sharon K., Paul W.B. Atkins, and Carolyn M. Axtell (2008), “Building Better Workplaces through Individual Perspective Taking: A Fresh Look at a Fundamental Human Process,” in *International Review of Industrial and Organizational Psychology*, ed. G. Hodgkinson and K. Ford, Wiley, 149–96.

Parker, Sharon K and Carolyn M. Axtell (2001), “Seeing Another Viewpoint : Antecedents and Outcomes of Employee Perspective Taking,” *Academy of Management Journal*, 44(6), 1085–1100.

Pennebaker, James W., Ryan L. Boyd, Kayla Jordan, and Kate Blackburn (2015), *The Development and Psychometric Properties of LIWC2015*, Austin, TX: University of Texas at Austin, http://hdl.handle.net/2152/31333.

Pryzant, Reid, Young joo Chung, and Dan Jurafsky (2017), “Predicting Sales from the Language of Product Descriptions,” *CEUR Workshop Proceedings*, 2311.

Rai, Arun (2020), “Explainable AI: From Black Box to Glass Box,” *Journal of the Academy of Marketing Science*, 48(1), 137–41.

Ribeiro, Marco Tulio, Sameer Singh, and Carlos Guestrin (2016), “‘Why Should i Trust You?’ Explaining the Predictions of Any Classifier,” *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 13-17-Augu, 1135–44.

Rocklage, Matthew D and Russell H Fazio (2020), “The Enhancing Versus Backfiring Effects of Positive Emotion in Consumer Reviews,” *Journal of Marketing Research*, Forthc., https://doi.org/10.1177/0022243719892594.

Rosen, Dennis L and Richard W Olshavsky (1987), “A Protocol Analysis of Brand Choice Strategies Involving Recommendations,” 14(3), 440–44.

Rousseau, Denise M., Sim B. Sitkin, Ronald S. Burt, and Colin Camerer (1998), “Not so Different After All: A Cross-Discipline View of Trust,” *Academy of Management Review*, 23(3), 393–404.

Rude, Stephanie S., Eva Maria Gortner, and James W. Pennebaker (2004), “Language Use of Depressed and Depression-Vulnerable College Students,” *Cognition and Emotion*, 18(8), 1121–33.

Schmitt, Bernd (2019), “From Atoms to Bits and Back: A Research Curation on Digital Technology and Agenda for Future Research,” *Journal of Consumer Research*, 46(4), 825–32.

Schmitt, Bernd H., June Cotte, Markus Giesler, Andrew T. Stephen, and Stacy Wood (2021), “Our Journal, Our Intellectual Home,” *Journal of Consumer Research*, 47, 633–35.

Schweitzer, Fiona, Russell Belk, Werner Jordan, and Melanie Ortner (2019), “Servant, Friend or Master? The Relationships Users Build with Voice-Controlled Smart Devices,” *Journal of Marketing Management*, 35(7–8), 693–715, https://doi.org/10.1080/0267257X.2019.1596970.

Seraj, Sarah, Kate G. Blackburn, and James W. Pennebaker (2021), “Language Left behind on Social Media Exposes the Emotional and Cognitive Costs of a Romantic Breakup,” *Proceedings of the National Academy of Sciences*, 118(7), https://www.pnas.org/content/118/7/e2017154118.

Settanni, Michele and Davide Marengo (2015), "Sharing Feelings Online: Studying Emotional Well-Being via Automated Text Analysis of Facebook Posts," *Frontiers in Psychology*, 6(JUL), 1–7.

Thomson, Matthew, Deborah J. MacInnis, and C. Whan Park (2005), "The Ties That Bind: Measuring the Strength of Consumers' Emotional Attachments to Brands," *Journal of Consumer Marketing*, 15(1), 77–91, http://www.emeraldinsight.com/10.1108/09590551211274928%5Cnhttp://psycnet.apa.org/journals/apl/77/6/975/%5Cnhttp://jvm.sagepub.com.proxy1.dom1.nhtv.nl/content/13/1/45%5Cnhttp://jvm.sagepub.com.proxy1.dom1.nhtv.nl/content/13/1/45.full.pdf%5Cnhttp://eprints.

Timoshenko, Artem and John R. Hauser (2019), "Identifying Customer Needs from User-Generated Content," *Marketing Science*, 38(1), 1–20, https://doi.org/10.1287/mksc.2018.1123.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin (2017), "Attention Is All You Need," in *Advances in Neural Information Processing Systems*, 5999–6009, https://arxiv.org/abs/1706.03762.

Wang, Alex, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy, and S. R. Bowman (2018), "GLUE: A Multi-Task Benchmark and Analysis Platform for Natural Language Understanding," *arXiv*, 1–20.

Waytz, Adam, John Cacioppo, and Nicholas Epley (2010), "Who Sees Human? The Stability and Importance of Individual Differences in Anthropomorphism," *Perspectives on Psychological Science*, 5(3), 219–32.

Waytz, Adam, Kurt Gray, Nicholas Epley, and Daniel M. Wegner (2010), "Causes and Consequences of Mind Perception," *Trends in Cognitive Sciences*, 14(8), 383–88, http://dx.doi.org/10.1016/j.tics.2010.05.006.

Waytz, Adam, Joy Heafner, and Nicholas Epley (2014), "The Mind in the Machine: Anthropomorphism Increases Trust in an Autonomous Vehicle," *Journal of Experimental Social Psychology*, 52, 113–17, http://dx.doi.org/10.1016/j.jesp.2014.01.005.

Waytz, Adam and Michael I. Norton (2014), "Botsourcing and Outsourcing: Robot, British, Chinese, and German Workers Are for Thinking-Not Feeling-Jobs," *Emotion*, 14(2), 434–44, https://doi.org/10.1037/a0036054.

Williams, Michele (2007), "Building Genuine Trust through Interpersonal Emotion Management : A Threat Regulation Model of Trust and Collaboration across Boundaries," *The Academy of Management Review*, 32(2), 595–621.

Wolf, Thomas (2020), "An Introduction to Transfer Learning in NLP and HuggingFace with Thomas Wolf," https://www.youtube.com/watch?v=t86G11tfVNw.

Yang, Linyun W, Pankaj Aggarwal, and Ann L McGill (2020), "The 3 C's of Anthropomorphism : Connection , Comprehension , and Competition," *Consumer Psychology Review*, 3(1), 3–19.

Zheng, Siqi, Jianghao Wang, Cong Sun, Xiaonan Zhang, and Matthew E. Kahn (2019), "Air Pollution Lowers Chinese Urbanites' Expressed Happiness on Social Media," *Nature Human Behaviour*, 3, 237–243, http://dx.doi.org/10.1038/s41562-018-0521-2.

# CURRICULUM VITAE

## PERSONAL INFORMATION

| | |
|---|---|
| Name | **Anouk Samantha Bergner (née Schneider)** |
| Date of Birth | **06.10.1984** |
| Place of Birth | **Biel, Switzerland** |

## EDUCATION

| | |
|---|---|
| 2018-Present | **PhD Candidate,** *TechX Lab, IBT-HSG, University of St.Gallen,* Switzerland<br>Advisor: Christian Hildebrand (PhD) |
| 2007-2008 | **MSc (distinction) in Decision Sciences,** *London School of Economics,* U.K.<br>Advisor: Barbara Fasolo (PhD) |
| 2003-2007 | **BA (magna cum laude) in Psychology,** *Princeton University* (U.S.A.)<br>Advisor: Danny Oppenheimer (PhD) |

## KEY RESEARCH INTERESTS

- Theory of Mind
- Computational Linguistics
- Technology-Augmented Decision Making

## JOURNAL PUBLICATIONS

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*.

**Bergner, A.**, (2020). Adaptive Sales Automation: Chatbots as Personalized and Scalable Sales Agents. *Marketing Review St. Gallen.*

**Bergner, A.**, Oppenheimer, D. M., & Detre, G. (2019). VAMP (Voting Agent Model of Preferences): A Computational Model of Individual Multi-attribute Choice. *Cognition.*

## SUBMISSIONS UNDER REVIEW & WORKING PAPERS

**Bergner, A.**, Hildebrand, C., & Häubl, G. Machine Talk: How Conversational Interfaces Promote Brand Intimacy and Influence Consumer Choice. *Journal of Consumer Research* (Revise & Resubmit; Third Round).

**Bergner, A.**, Hartmann, J., & Hildebrand, C. Conferring Minds to Machines: A Deep Learning Approach to Mind Perception, Smart Object Relationships, and Task Delegation (In Preparation for Submission)

**CONFERENCE PROCEEDINGS & ORAL PRESENTATIONS**

**Bergner, A.**, Hartmann, J., & Hildebrand, C. (2021). Conferring Minds to Machines: A Deep Learning Approach to Mind Perception, Smart Object Relationships, and Task Delegation. *Artificial Intelligence in Management Conference (virtual).*

**Bergner, A.**, Hartmann, J., & Hildebrand, C. (2021). Conferring Minds to Machines: A Deep Learning Approach to Mind Perception, Smart Object Relationships, and Task Delegation. *European Marketing Association Conference (virtual).*

**Bergner, A.**, Hartmann, J., & Hildebrand, C. (2021). Conferring Minds to Machines: A Deep Learning Approach to Mind Perception, Trust, and Task Delegation. *Society for Consumer Psychology Conference (virtual).*

**Bergner, A.**, Hartmann, J., & Hildebrand, C. (2020). Conferring Minds to Machines: A Deep Learning Approach to Mind Perception, Technology Attachment, and Trust. *Association of Consumer Research Conference, Paris (virtual).*

**Bergner, A.**, Hildebrand, C. & Häubl, G. (2020). Machine Talk: How Conversational Interfaces Promote Brand Intimacy and Influence Consumer Choice. *European Marketing Association Conference, Budapest, Hungary*.

**Bergner, A.**, Hildebrand, C., & Häubl, G. (2019). Conversational Interfaces as Persuasion Instruments: Implications for Consumer Choice and Brand Perceptions (Special Session on "The Machine Age of Consumer Research: How Robot-Based Expression Modalities Alter Perception and Choice"), *Association for Consumer Research Conference, Atlanta, USA.*

**Bergner, A.**, Hildebrand, C., & Häubl, G. (2019). Machine Talk: How Conversational Interfaces Promote Brand Intimacy and Influence Consumer Choice, *Society for Consumer Psychology, Savannah, USA.*

**Bergner, A.**, Hildebrand, C., & Häubl, G. (2019). Machine Talk: How Conversational Interfaces Promote Brand Intimacy and Influence Consumer Choice. *American Marketing Association CBSIG, Berne, Switzerland.*

Hildebrand, C., **Bergner, A.**, & Häubl, G. (2019). Machine Talk: How Conversational Interfaces Promote Brand Intimacy and Influence Consumer Choice. *Theory and Practice in Marketing Conference, New York, USA.*

Hildebrand, C., & **Bergner, A.** (2019). Detrimental Trust in Automation: How Conversational Robo Advisors Leverage Trust and Miscalibrated Risk Taking (Special Session on "Persuasion Dynamics of Technology-Mediated Social Presence in Consumer-Firm Interactions"), *Association for Consumer Research Conference, Atlanta, USA.*

Hildebrand, C., & **Bergner, A.** (2019). Detrimental Trust in Automation: How Conversational Robo Advisors Leverage Trust and Miscalibrated Risk Taking, *Society for Consumer Psychology, Savannah, USA.*

**Bergner, A.** (2019). Conversational Interfaces: A Powerful Tool for Shaping Consumption Experiences, Influencing Preferences, and Altering Decision-Making, *EMAC Doctoral Colloquium, Hamburg, Germany.*

**Bergner, A.**, Hildebrand, C., & Häubl, G. (2019). Machine Talk: How Conversational Interfaces Promote Brand Intimacy and Influence Consumer Choice, *Rohrschacher Kreis, The Future of Marketing, Luzern, Switzerland.*

**Bergner, A.**, Hildebrand, C., & Häubl, G. (2019). Machine Talk: How Conversational Interfaces Promote Brand Intimacy and Influence Consumer Choice, *Forschungstagung Marketing, Berlin, Germany.*

**Bergner (Schneider), A.,** & Oppenheimer, D. M. (2007). Application of voting geometry to multialternative choice. In *Proc. 29th Annu. Meet. Cogn. Sci. Soc.*, pp. 635–40.

## SERVICE

**Ad hoc Reviewer**, Management Science; Journal of Business Research

## TEACHING EXPERIENCE

2019-2021 **Teaching Assistant,** *University of St. Gallen, St. Gallen*
Judgement and Decision-Making, Consumer Behavior & Methods

2017-2018 **Lecturer,** *Haute école de gestion (HEG), Geneva*
Full-time lecturer in Marketing for Bachelor students

## PROFESSIONAL EXPERIENCE

2016-2017 **Director**
Consumer & Market Insights, *COTY Global Professional Beauty, Geneva*

2015-2016 **Senior Manager**
Retail Strategy & Shopper Insights. *Procter & Gamble, Global Luxury, Geneva*

2012-2015 **Manager**
Retail Strategy & Shopper Insights, *Procter & Gamble, WE Beauty & Grooming, London*

2010-2012 **Manager**
Consumer & Market Insights, *Procter & Gamble, Global Feminine Care, Geneva*

2008-2010 **Associate Manager**
Consumer & Market Insights, *Procter & Gamble, Feminine Care Western Europe*

## ACADEMIC AWARDS & HONORS

2019 **Best Paper Finalist** *Society for Consumer Psychology (SCP)* for Bergner, A., Hildebrand, C., & Häubl, G. (2019). Machine Talk: How Conversational Interfaces Promote Brand Intimacy and Influence Consumer Choice

2008 **Award for best Dissertation** in Decision Sciences, *London School of Economics*

2007 **Hillel Einhorn New Investigator Award,** *Princeton University*

2007 **George A. Miller Prize** for the best interdisciplinary thesis, *Princeton University*