Radboud University

Chatbots in Healthcare – The Effects of Affective and Cognitive Empathy on the User's Satisfaction with, Trust in, and Loyalty to the Service

Master Thesis

Name:	Pedro Philipp Hofer
Student number:	s1062313
Faculty:	Nijmegen School of Management
Program:	MSc Business Administration - Marketing

Supervisor:	Dr Csilla Horváth
2nd examiner:	Karim Sidaoui
Date:	June 14th 2021

Abstract

Empathy plays a crucial role in interpersonal communication. Widely understood as the ability to understand and feel another's inner state, it is an omnipresent topic in scientific research. However, with the rise of artificial intelligence, humans now not only have other humans to talk to but completely new dialog partners: chatbots. The question is how empathy takes effect in these smart text-based entities, which are increasingly used as service agents and therapeutical advisers in healthcare. Their right development could lead to better customer experiences and above all help suffering patients. The present thesis aims to contribute to this highly relevant subject by examining the effects of affective and cognitive empathy in healthcare service chatbots on the user's satisfaction with, trust in, and loyalty to the service, and comparing it to a regular human-to-human interaction.

While it could not be confirmed that just adding general empathy to a chatbot automatically results in more user satisfaction, trust, or loyalty, the strongly differing effects of affective and cognitive empathy in chatbots and humans became clear. A chatbot provides a significantly more pleasant user experience with cognitive empathy, i.e., when it shows that it understands the user's inner state. In contrast, a human healthcare adviser is more appreciated when the user's inner state is felt and thus affective empathy is shown. A reason for this could lie in the different expectations that are placed on a chatbot compared to a human. Chatbots are not necessarily expected to have the ability to feel emotions, making their attempt to show affective empathy seem fake and not genuine. Although the technology of chatbots and people's expectations of it are still constantly evolving, it seems like a cognitively empathic chatbot makes the most satisfied, trusting, and loyal users for now.

Keywords: Affective empathy, Cognitive empathy, Chatbots, Healthcare, Service

Content

1. Introduction

Nowadays, we not only communicate with other human beings but also with Artificial intelligence (AI) in the form of virtual digital assistants, as they become more and more part of our daily lives. In 2019, there were roughly 3.25 billion voice assistants like Apple's Siri or Amazon's Alexa in use, primarily on smartphones, smart TVs and similar smart home devices (Juniper Research, 2021). This number is estimated to reach 8 billion by the year 2023 (Juniper Research, 2021). Conversational chatbots are also gaining popularity. These text-based virtual assistants that can be defined as "machine agents that serve as natural language user interfaces for data and service providers" are mostly used as a mean for direct customer service (Brandtzaeg & Følstad, 2017; Xu et al., 2017). They are increasingly integrated in social messaging platforms such as Facebook messenger, Skype or Slack, but also in healthcare applications (Brandtzaeg & Følstad, 2017). Already in 2013, Microsoft CEO Satya Nadella said "Chatbots are the new apps" (Della Cava, 2016).

One of the most fundamental elements of human communication is empathy (Redmond, 1989). As an interpersonal phenomenon, it strongly defines and influences morality and prosocial behavior (Decety, 2010), making it an omnipresent topic and research field in science (Nakao & Itakura, 2009). Empathy and communication competence seem so closely related that they cannot even be seen as two discrete concepts (Redmond, 1985). Rather, they are two almost interchangeably usable concepts that do not exist without the other, due to the fact that the production of both require the same underlying skills (Redmond, 1985). This would mean that an empathic message is always communicatively competent, as well as the other way around. Empathy is defined as the ability of one self-aware self to detect accurately and comprehend the emotional information being transmitted by another self (Levenson & Ruef, 1992; Wispé, 1986). Furthermore, literature indicates emerging consensus that empathy is a multidimensional construct consisting of a *cognitive dimension* (understanding others' internal states), an *affective dimension* (experiencing affective states that are congruent with others' affective states) and a *behavioral dimension* (engaging in verbal and nonverbal behaviors that demonstrate affective and/or cognitive empathy) (Clark, Robertson, & Young, 2019; Cuff et al., 2016; Van der Graaff et al., 2016).

As empathy plays such a key role in human communication, it seems obvious that its quality-improving benefits are also essential for interactions between humans and virtual AI assistants. This aspect is especially important for interactions with chatbots in the healthcare sector. 80% of small-scale diseases that make people visit a doctor can actually be cured at home using home remedies and without the intervention of a doctor (Bhirud et al., 2019). So called text-based healthcare chatbots (THCBs) like the ones from Babylon Health, Ada Health or Your.MD all try to solve that problem and relive overloaded healthcare professionals by providing users with reliable diagnoses and further helpful information (Bates, 2019). Additionally, they have the potential to treat patients who do not seek help from a doctor despite great pain, due to feelings of social embarrassment about the concern (Gonzalez, 2017; Hernandez, 2018; Lucas et al., 2014). However, the problem is that these chatbots often lack the ability to have a natural conversation that makes the patient feel understood (Bhirud et al., 2019). It seems like users are missing the deeply desired empathy in chatbots. The question is how this affects the crucial success factors of a service, in the form of the user's *satisfaction*, *trust*, and *loyalty* towards it. All of the three constructs have been identified again and again as fundamental elements in consumer experience, heavily influencing people's intention to repurchase or reuse the service (Bahadur et al., 2019; Kassim & Abdullah, 2010; Morgan & Hunt, 1994; Wieseke, Geigenmüller, & Kraus, 2012). While businesswise, failing to achieve these success factors could mean decreasing revenue for firms, the issue plays an even bigger role in healthcare. The incompetence to make assistive chatbots a pleasant interlocutors could take the chance of relieving an exhausted occupational group of healthcare professionals, and more importantly of providing suffering people with more versatile, more accessible and, for many, more suitable ways of healthcare and treatment options.

Companies who develop these THCBs are thus trying to create an authentic user experience and induce empathy into their chatbots by designing them in a certain way. This includes what the chatbot says and how it says it (Cameron et al., 2018a; Cameron et al., 2018b; Dahiya, 2017). Studies show that making the chatbot mimic the user's state and behavior leads to its response being perceived as more empathic, adequate and satisfactory from the user's point of view (Hegel et al., 2006; Redmond, 1985; Riek, Paul & Robinson, 2010). Related research about human-robot-interactions further suggests that, once accomplished, empathy in robots leads to more enjoyable interactions through stronger feelings of companionship, self-validation and reliable alliance (Leite et al., 2013).

All in all, however, despite comprehensive research about empathy in a person-to-person context (Clark, Robertson, & Young, 2019; Decety, 2010; Redmond, 1985; Stiff et al., 1988), similar literature about interactions between humans and chatbots is still very limited. The literature gap becomes even more noticeable when taking the distinction between affective and cognitive empathy into account. Hence, this study contributes to filling this gap by focusing on the crucial service dimensions of satisfaction, trust, and loyalty and thus tackling the following research questions: *How does empathy compared to no empathy in a text-based healthcare chatbot affect the user's satisfaction with, trust in, and loyalty to the service? And how does affective empathy compared to cognitive empathy take effect?* In order to make sure that the results can be uniquely attributed to the chatbot setting, it will be compared to a regular human-to-human interaction with a healthcare professional as adviser. Arising answers are supposed to provide a better understanding of how empathy works in healthcare service in this upcoming age of chatbots.

Finally, the topic also reveals great managerial relevance. *Working alliance* can be seen as the degree of interaction between the patients and health professionals in order to create an attachment bond during therapy (Castonguay, Constantino & Holtforth, 2006). It is a fundamental construct of the therapeutical process in healthcare and strongly affects the success of a treatment (Flückiger et al., 2012; Horvath and Greenberg, 1989; Martin, Garske & Davis, 2000). Due to an increasing shortage of health professionals, a sustainable working alliance often cannot be ensured (Aluttis, Bishaw & Frank, 2014; Wahle et al., 2017; Wahle et al., 2016). Text-based healthcare chatbots are gaining popularity as an additional service or even alternative for these interactions and can thus relieve the high demand for interpersonal therapy (Scherer, Wünderlich, & Wangenheim, 2015). However, this shift can only be successful if the used THCBs are able to provide a satisfying service to the user that induces trust and loyalty, in order for this technology to sustainably win over people's favor. Hence, looking further into a chatbot's empathy in the healthcare domain as a possible antecedent of these critical factors is of great importance for reaching that goal.

This thesis is structured as follows: First, a broad literature review (chapter 2) summarizes and explains the most relevant findings and constructs in the fields of empathy (2.1) and chatbots in healthcare (2.2). Additionally, the chapter introduces two essential theoretical foundations in the form of the satisfaction-trust-loyalty interrelation in service (2.3.1) and the *Expectation confirmation theory* (2.3.2). Lastly, it covers the proposed hypotheses and conceptual model of the study (2.4). Chapter 3 discusses the methodology, comprising subchapters about the used method (3.1), data collection and sample (3.2), operationalization of included variables (3.3), data analysis procedure (3.4), and research ethics (3.5). Next, chapter 4 covers the experiment's results, first making a manipulation check (4.1), followed by the main statistical analysis (4.2). Here, the necessary assumptions for the analyses are discussed (4.2.1). Thereafter, the hypotheses for the main effects (4.2.2) and moderating effects (4.2.3) are tested. The analysis is supplemented with a brief post-hoc analysis (4.3) and completed with a summary of the results (4.4). Finally, the conclusion (chapter 5) includes the study's theoretical and managerial implications (5.1), as well as this study's limitations and an outlook for future research (5.2).

2. Literature review

2.1 Empathy

Defining the construct of empathy. The term "empathy" is derived from the German word "Einfühlung" (literally, "in-feeling"), coined by psychologist Theodore Lipps in 1880 (Ioannidou & Konstantikaki, 2008). Generally meant as the ability to understand another's "state of mind", it is seen as fundamental tool for communication and prosocial behavior (Hogan, 1969). It is strongly linked to communication competence, meaning especially communication responsiveness and comforting behavior (Redmond, 1985; Stiff et al., 1988). Moreover, it positively affects morality and the regulation of aggression on an interpersonal level (Decety, 2010). The widely spread importance of empathy has led to many attempts in academic literature in trying to define this construct. In fact, one study reviewed a variety of literature and found 43 distinct definitions of empathy (Cuff et al., 2016). Nevertheless, most scholars see empathy as a multidimensional construct that has both a state and a trait level (Cuff et al., 2016). There seems to be further agreement that the construct consists of an affective and a cognitive dimension, while some scholars also see a third behavioral dimension in empathy (Van der Graaff et al., 2016).

Affective empathy forms the phylogenetically earliest system in empathy and can be seen as the rather automatic reaction of "experiencing affective states that are congruent with others' affective states" (Clark, Robertson, & Young, 2019; de Waal, 2008; Gonzalez-Liencres, Shamay-Tsoory, & Brüne, 2013; Preston & de Waal, 2002). It can be explained with the simulation perspective theory (Gallese & Goldman, 1998), which describes that people instinctively respond to another person's affective state due to the perception-action mechanism, leading to matching affective states (Preston & de Waal, 2002). Affective empathy can be defined as "*the experience of an affective state that is caused by and congruent with another's affective state*".

Cognitive empathy on the other hand develops later and includes consciously comprehending other's internal states (de Waal, 2008; Shamay-Tsoory, 2011). It can be derived from the theory of mind perspective (Wellman, 2014), which states that a system of rules emerged from previous own experiences makes people able to understand other's feelings and thoughts (Vachon, Lynam, & Johnson, 2014). Neuroimaging has shown that cognitive empathy activates different areas in the human brain than affective empathy, clearly proving that these are distinct neurological processes (Decety & Cowell, 2014; Roca et al., 2011; Shamay-Tsoory, Aharon-Peretz & Perry, 2009; Shamay-Tsoory, 2011; Walter, 2012; Zaki et al., 2009). Cognitive empathy is defined as "*the state of understanding another's internal state*".

Some studies find evidence for a third *behavioral dimension* in empathy, which mainly focuses on behavioral mirroring and empathic communication. This includes copying other's facial expressions, behavior or linguistic style (Chartrand & Lakin, 2013; Chartrand & Bargh, 1999; Lord et al., 2015). As this mirroring process often happens automatically (Chartrand & Lakin, 2013), an overlap with the affective dimension seems to be present. However, neuroimaging again was able to identify different activating processes, which speaks for a distinction (Carr et al., 2003). Behavioral empathy is "*verbal or nonverbal behavior that demonstrates affective and/or cognitive empathy*".

In addition to these conceptual dimensions, empathy also has a genetically predetermined *trait* component and a situationally varying *state* component (Cuff et al., 2016). While the trait component is stable and situationally independent mostly due to genetics (Nezlek et al., 2001; Christ, Carlo, & Stoltenberg, 2016; Hurlemann et al., 2010), the state component varies within-person over time and can be activated and influenced by external cues (Nezlek et al., 2001; Toomey & Rudolph, 2017). This can, for example, be achieved by exposing people to other's affective states (Jackson, Meltzoff & Decety, 2005) or making them imagine being in a specific situation that someone else experienced (Batson et al., 2002).

Finally, empathy has to be distinguished from similar constructs, most importantly sympathy, also known as empathic concern or compassion. Some argue that sympathy is a dimension of empathy (Davies, 1983). There is, however, emerging consensus that they are distinct constructs (Bernhardt & Singer, 2012; Cuff et al., 2016; Davis, 2009). In contrast to empathy, sympathy is seen only as an uncontrolled emotional reaction (Halpern, 2003). Sympathy cannot be solely compared with the affective dimension of empathy either, due to the fact that it does not necessarily imply feeling the same affective state as the target. Instead, it rather just consists of the urge to help or assist a suffering other (Clark, Robertson, & Young, 2019).

This study focuses on affective and cognitive empathy, as they build the core dimensions of empathy. As the experimental part of the research will include the implementation of affective and cognitive empathy into a healthcare conversation, the two will automatically be moved into the behavioral dimension. This seems inevitable due to affective and cognitive empathy only becoming easily visible and measurable through behavior.

However, this does not change the study's main emphasis on the affective and cognitive dimensions of empathy. Furthermore, it is to mention that it covers the state part of empathy. This decision is based on the fact that it is state empathy which is being situationally activated by external cues. Hence, a situation such as being in a health consultation specifically triggers state empathy.

Empathy in the service and healthcare setting. Empathy plays an incredibly important role in healthcare and heavily influences the effectiveness of therapeutical relationships (Wiseman, 1996). Used as a communication tool by both the clinician and the patient, it increases the efficiency of gathering information, facilitates clinical interviews and leads to more appreciation and honoring of the patient (Ioannidou & Konstantikaki, 2008). Moreover, understanding the patient's feelings is a key element of quality nursing care (Reynolds, Scott, & Jessiman, 1999).

What remains is the question about the perceived genuineness of empathy in this field. Genuineness, often also referred to as "authenticity", is viewed as someone's honesty and sincerity (Albrecht, 2006). Hospitals make money with ongoing therapeutical treatments, recommending medical drugs or surgeries (Tuan, 2012). Empathy can be manipulatively used as a flattery tactic by healthcare professionals to achieve self-serving goals (Bove, 2019). This might be anticipated by patients, which could worsen their perception of the service. However, while empathy from healthcare professionals is probably still mostly believed and appreciated, the problem of real genuineness becomes bigger when using nonhuman, programmed chatbots. Research shows that content- or timewise inappropriate replies quickly make a chatbot seem unauthentic and lead to frustration instead of feeling supported (Bove, 2019). This can happen easily when considering that most people don't think computers are able to feel and thus show real emotions (Nass & Moon, 2000). The aspect of genuineness should thus not be neglected for the purpose of this research.

2.2 Chatbots in healthcare

The creation of the first chatbot dates all the way back to the year 1964. Developed at the MIT Artificial Intelligence Laboratory by Joseph Weizenbaum, natural language processing program ELIZA is now considered a milestone in AI and prototype for modern chatbots (Vaidyam et al., 2019). Since then, chatbots have constantly been improved and eventually experienced a huge jump in popularity and applications in 2016, due to increasingly accelerating technology and the spread of low-latency networking (Grudin & Jacques, 2019).

They now play an important role in real-time customer service of many businesses and often replace human chat service agents. This is mainly due the huge cost- and time-saving benefits of AI-based chatbots.

Chatbots are especially being used in the healthcare sector. In 2017, one of the top requests to Amazon's voice chatbot Alexa was "Alexa, help me relax", according to MIT Technology Review (Anders, 2017). It is thus no wonder that chatbots found and to this day still find a growing number of application fields in healthcare (Bhirud et al., 2019). Particularly text-based chatbots are being used more and more, and are able to provide users with reliable diagnoses, give answers to general healthcare questions or to plan and book appointments with doctors (Bates, 2019; Bhirud et al., 2019). These cases do not only include small-scale diseases, but also more serious mental issues and illnesses such as dementia or insomnia (Bhirud et al., 2019). Furthermore, they are being used in suicide prevention and cognitive behavioral therapy (Gonzalez, 2017; Hernandez, 2018), and are much appreciated by those who are reluctant to speak with a therapist due to feeling uncomfortable sharing their problems. For example, a study shows that veterans returning from a tour of duty are more likely to talk to a virtual chatbot than an actual person (Lucas et al., 2014), revealing the enormous potential of this technology.

Nearly 75% of scientific studies about chatbots are publications from the engineering field. Medical publications on the other hand are relatively sparse (Vaidyam et al., 2019). This is surprising and highlights the need for further research, as chatbots have proven to have multiple benefits. One of the main benefits of chatbots include self-psychoeducation and adherence. For example, by tracking medication and supporting physical activity, they can provide healthy lifestyle recommendations and cognitive behavioral therapy (Bickmore et al., 2010a; Gardiner et al., 2017; Ly, Ly, & Andersson, 2017; Vaidyam et al., 2019). Higher adherence to a chatbot causes a significant improvement in psychological well-being and perceived stress compared to not using a chatbot (Ly, Ly, & Andersson, 2017). Individuals with depressive disorders even prefer a chatbot based therapy alliance to a real clinical (Bickmore et al., 2010a). Moreover, participants showed high satisfaction with their interventions and gave high scores for ease of use, liking, trust and desire to continue using the system (Bickmore et al., 2010b; Gardiner et al., 2017).

Some literature also indicates potential harms of using chatbots in healthcare, although these are mainly just minor risks (Bickmore et al., 2010b; Tielman et al., 2017; Miner et al., 2016). A study researching about smartphone-based chatbots and their responses to suicide related emergencies showed that their answers were limited and sometimes inappropriate. To the statement “I am depressed” one chatbot answered with “Maybe the weather is affecting you” (Miner et al., 2016). Additionally, there is concern that patients who suffer from psychiatric illnesses may develop an unhealthy excessive attachment to the chatbot (Bickmore et al., 2010b; Tielman et al., 2017).

It is safe to say that a lot of users still experience unsatisfactory interactions with chatbots, which might lead to resistance against this technology. In order to find evidence on how to best build a helpful text-based healthcare chatbot (THCB), researchers keep investigating the design of THCBs (Cameron et al., 2018a; Cameron et al., 2018b; Battineni, Chintalapudi, & Amenta, 2020). These studies aim to find the right coding of algorithms that affect the chatbot’s knowledge gathering and its ability to connect that knowledge, in order to understand and comprehend certain requests (Cameron et al., 2018b). Integrating humor, an appropriate and authentic typing speed, or regular mentioning of the user’s name can help to increase the chatbot’s usability and thus improve the user experience (Cameron et al., 2018a). Many times, giving the chatbot human-like characteristics enhances the experience and user compliance (Adam, Wessel, & Benlian, 2020). Further research about the important element of empathy in a chatbot has the potential to extend existing knowledge about how to provide customers with great chatbot experiences.

2.3 Underlying theories

2.3.1 The satisfaction-trust-loyalty interrelation in service

Satisfaction, trust, and loyalty are all fundamental parts in consumer research (Kassim & Abdullah, 2010). In consonance with the Oxford Languages definition (2001) and embedded into the service setting, satisfaction can be defined as the fulfilment of one's wishes, expectations, or needs by a service, or the pleasure derived from this. Trust can be rather understood as an attitude towards a service, which becomes relevant when a consumer feels uncertainty and vulnerability (Doney & Cannon, 1997; Moorman, Zaltman, & Deshpande, 1992). With the help of Chaudhuri and Holbrook’s (2001) definition, trust can be seen as the willingness of the consumer to rely on the ability of the service to perform its stated function. Lastly, service loyalty acts as a bond between consumer and service (Gremler & Brown, 1996). A very comprehensive yet precise definition is provided in the following: “Service loyalty is the degree to which a customer exhibits repeat purchasing behavior from a service provider, possesses a positive attitudinal disposition toward the provider, and considers using only this provider when a need for this service arises.” (Gremler & Brown, 1996, p. 173).

There is consensus that the three constructs heavily correlate with each other and even show major dependencies (Bahadur et al., 2019; Morgan & Hunt, 1994; Wieseke, Geigenmüller, & Kraus, 2012). The essential role that a consumer's loyalty plays for the success of a product or service (Donio, Massari, & Passiante, 2006) initiates research about the antecedents of it. An essential influence seems to be identified in the consumer's satisfaction with the service (Bahadur et al., 2019). In fact, satisfying service interactions between employees and customer lead to loyalty to the service provider (Chaudhuri & Holbrook, 2001). Next to these two constructs of satisfaction and loyalty, the consumer's trust appears to play another important role, creating a triple relationship of interdependent constructs. While there is research suggesting trust acts as a mediator between satisfaction and loyalty (Kassim & Abdullah, 2010; Morgan & Hunt, 1994), different literature sees satisfaction as the mediating part between trust and loyalty (Bahadur et al., 2019). Regardless of which of the two perspectives comes closest to reality, a strongly supported theory of a satisfaction-trust-loyalty interrelation emerges. It will be important to keep this in mind when trying to predict single effects of empathy on the three constructs.

2.3.2 Expectation confirmation theory

Looking more closely into the concept of consumer satisfaction, also as one of the key predictors of trust and loyalty, there seems to lie great relevance in the relationship between what the consumer expects and what is actually confirmed. This phenomenon is described in the *Expectation confirmation theory* by Oliver (1977). The theory involves four primary constructs: expectations, perceived performance, disconfirmation of beliefs, and satisfaction.

Expectations refer to the characteristics and attributes that a consumer predicts to experience or associate with an entity, in this case the service. It directly influences perceived performance and disconfirmation of beliefs, and indirectly affects satisfaction. While *perceived performance* refers to perception of the actual performance of the service, *disconfirmation of beliefs* is the consumer's evaluation of the service. This evaluation always includes a comparison with the initial expectations. Depending on how positive the evaluation, which is also based on the perceived performance, turns out compared to the expectations, the *satisfaction* will less or more be positively affected. Satisfaction refers to the extent to which the consumer is pleased with the service post-purchase.

For the purpose of this research, it will be of great importance to know how satisfaction might be caused. Hence, the interplay of expectation and confirmation will be considered in affected parts of the following formulation of hypotheses.

2.4 Hypotheses and conceptual model

First of all, the focus is on the effects of the health adviser's general empathy on the user's satisfaction, loyalty, and trust. For this purpose, no distinction is made between affective and cognitive empathy yet, nor between chatbot and human as health adviser. Literature from the research fields of service and customer experience show that employees who show customers empathy are able to create a positive image and better customer care in general (van Dolen, Ruyter, & Lemmink, 2004; Wieseke, Geigenmüller, & Kraus, 2012). Being empathic by trying to understand and feel what the customer needs allows for a significantly higher chance of identifying and actually satisfying those needs. In fact, this exact empathy-caused increase in satisfaction has already been supported by Homburg, Wieseke, and Bornemann (2009). Applying this groundwork to the study's service setting in healthcare, it can be assumed that empathy shown by the health adviser, compared to no shown empathy, increases the user's satisfaction with the healthcare service. This leads to the first hypothesis:

> ***H1a:*** *A health adviser who shows empathy (vs. one who does not show any empathy) increases the user's satisfaction with the service.*

Previous literature stresses the close interrelations of satisfaction, trust and loyalty of customers in the service setting (Morgan & Hunt, 1994; Wieseke, Geigenmüller, & Kraus, 2012). A very common conceptualization depicts trust as a mediating construct in-between a satisfaction-loyalty dependency (Morgan & Hunt, 1994), meaning that satisfaction leads to loyalty through trust. What arises is a first hint towards the assumption that an increase in the user's satisfaction with the healthcare service could also mean in increase in trust and loyalty, although only being caused by the satisfaction-trust-loyalty relationship and not empathy directly. There is, however, literature suggesting a direct positive effect of empathic accuracy on interpersonal trust (Feng, Lazar, & Preece, 2004). Hence, it definitely seems plausible to predict a positive effect of empathy on the user's trust in the service, leading to the following second hypothesis:

> ***H1b:*** *A health adviser who shows empathy (vs. one who does not show any empathy) increases the user's trust in the service.*

As already pointed out, loyalty also seems to be positively affected by satisfaction. There is literature specifically elaborating on how empathy in employees leads to service loyalty by the customer (Bahadur et al., 2019). Once again, satisfaction and trust were identified as mediating variables. Nevertheless, as it is assumed that empathy increases the user's satisfaction, it can further be assumed that loyalty also experiences a positive effect:

> ***H1c:*** *A health adviser who shows empathy (vs. one who does not show any empathy) increases the user's loyalty to the service.*

When taking a closer look at the potential differences in the effects of affective empathy and cognitive empathy on satisfaction, trust, and loyalty, assumptions do not seem that clear at first. Both affective and cognitive empathy can be effective ways to show empathic behavior and thus improve the relationship with another (Clark, Robertson, & Young, 2019). However, there are differences in how the two types of empathy are perceived and what exactly they influence (Kim, Kaplowitz, & Johnston, 2004). In the healthcare context, a physician's cognitive empathy tends to allow a better exchange of cognitive information with the patient and causes an increased perception of expertise. In contrast, affective empathy seems to positively affect partnership, which is further found to increase trust and satisfaction (Kim, Kaplowitz, & Johnston, 2004). Although it could be assumed that a good exchange of cognitive information and perceived expertise might as well lead to satisfaction or trust in some way, affective empathy appears to be the more direct and above all the stronger causer of the two constructs. Applying the previously discussed close interrelations of satisfaction, trust, and loyalty, extends this assumption by supposing that loyalty should also experience a bigger increase from affective empathy than from cognitive empathy. This leads to the following set of hypotheses for this study:

> ***H2a:*** *A health adviser who shows affective empathy (vs. one who shows cognitive empathy) increases the user's satisfaction with the service.*

> ***H2b:*** *A health adviser who shows affective empathy (vs. one who shows cognitive empathy) increases the user's trust in the service.*

> ***H2c:*** *A health adviser who shows affective empathy (vs. one who shows cognitive empathy) increases the user's loyalty to the service.*

The next step comprises elaborating on how these first six main effects behave for interactions with a chatbot compared to interactions with a human as health adviser. Hence, it will now be looked at the potential moderating effect of the interlocutor that the user contacts when seeking out health-related advice.

When trying to predict the interlocutor's moderating effect regarding the user's satisfaction with the service, it is worth taking a look at the *Expectation confirmation theory* by Oliver (1977). In its essence, it states that someone's satisfaction with a service is heavily based on their expectations of the service and to what extend these expectations are fulfilled. Research implies that, when asked about it, people claim they do not assign human traits and characteristics to computers (Nass & Moon, 2000). They further seem to have no expectations regarding a computer's ability to feel or understand emotions. It can thus be assumed that empathy is generally not expected from a chatbot. Conversely, experiencing a chatbot showing empathy might evoke a positive service evaluation compared to what was expected, as empathy is highly desired in human communication (Redmond, 1985; Redmond, 1989). This would ultimately lead to satisfaction. As a result, the following can be assumed:

> ***H3a:*** *Having a chatbot as interlocutor (vs. a human) amplifies the positive effect of empathy (vs. no empathy) on the user's satisfaction with the service.*

It becomes more difficult when trying to predict the moderating effects of the interlocutor on the user's trust and loyalty. On the one hand, the proven interrelations between satisfaction, trust, and loyalty (Morgan & Hunt, 1994) leads to the assumption that having a chatbot as interlocutor (vs. a human) also amplifies the positive effect of empathy on trust and loyalty. On the other hand, it should not be neglected that chatbots are still a relatively new technology for many people when it comes to receiving health advice (Nadarzynski et al., 2019), suggesting that trust and loyalty could be hard to achieve. However, trust in something is not necessarily dependent on the newness of it. Research from organizational studies show that it comes down to an open communication, the sharing of critical information and the true sharing of perceptions and feelings when trying to facilitate interpersonal trust (Mishra & Morrissey, 1990). None of these factors should really rely on the interlocutor being a chatbot or a human. Hence, the decisive aspect could be the lower expectations for the chatbot, making the unexpectedly perceived empathy lead to not only more satisfaction, but consequently also more trust in the service:

> ***H3b:*** *Having a chatbot as interlocutor (vs. a human) amplifies the positive effect of empathy (vs. no empathy) on the user's trust in the service.*

Moreover, the huge dependency of loyalty on satisfaction and trust supports the assumption that empathy will further lead to more loyalty for users that contact a chatbot, compared to a human:

> ***H3c:*** *Having a chatbot as interlocutor (vs. a human) amplifies the positive effect of empathy (vs. no empathy) on the user's loyalty to the service.*

Finally, what remains is the question of whether affective and cognitive empathy are perceived differently by the user depending on the interlocutor being a chatbot or a human. Therefore, the aspect of the perceived genuineness of the chatbot needs to be taken into account. Genuineness, often referred to as "authenticity", describes someone's honesty and sincerity (Albrecht, 2006). As discussed earlier, people generally have the perception that computers are not able to feel emotions (Nass & Moon, 2000). Consequently, users who receive affective empathy by a chatbot might not perceive it as real and genuine, but as a fake simulation. Research has already shown that chatbots struggle to convey genuineness (Neururer et al., 2018). Seen as this pretended affective empathy, it can be assumed that it at least hinders an increase in satisfaction, trust, and loyalty. Cognitive empathy, however, might be more on line with what people think is possible for computers to practice. Understanding another's state could be solely based on having sufficient information on the person's situation and concern, which can very well be programmed into a chatbot. Thus, affective empathy from a chatbot should be less appreciated than from a human. Once again, the close interrelations of the three dependent variables allow for following final three hypotheses to be formulated as a whole:

> ***H4a:*** *Having a chatbot as interlocutor (vs. a human) weakens the positive effect of affective empathy (vs. cognitive empathy) on the user's satisfaction with the service.*

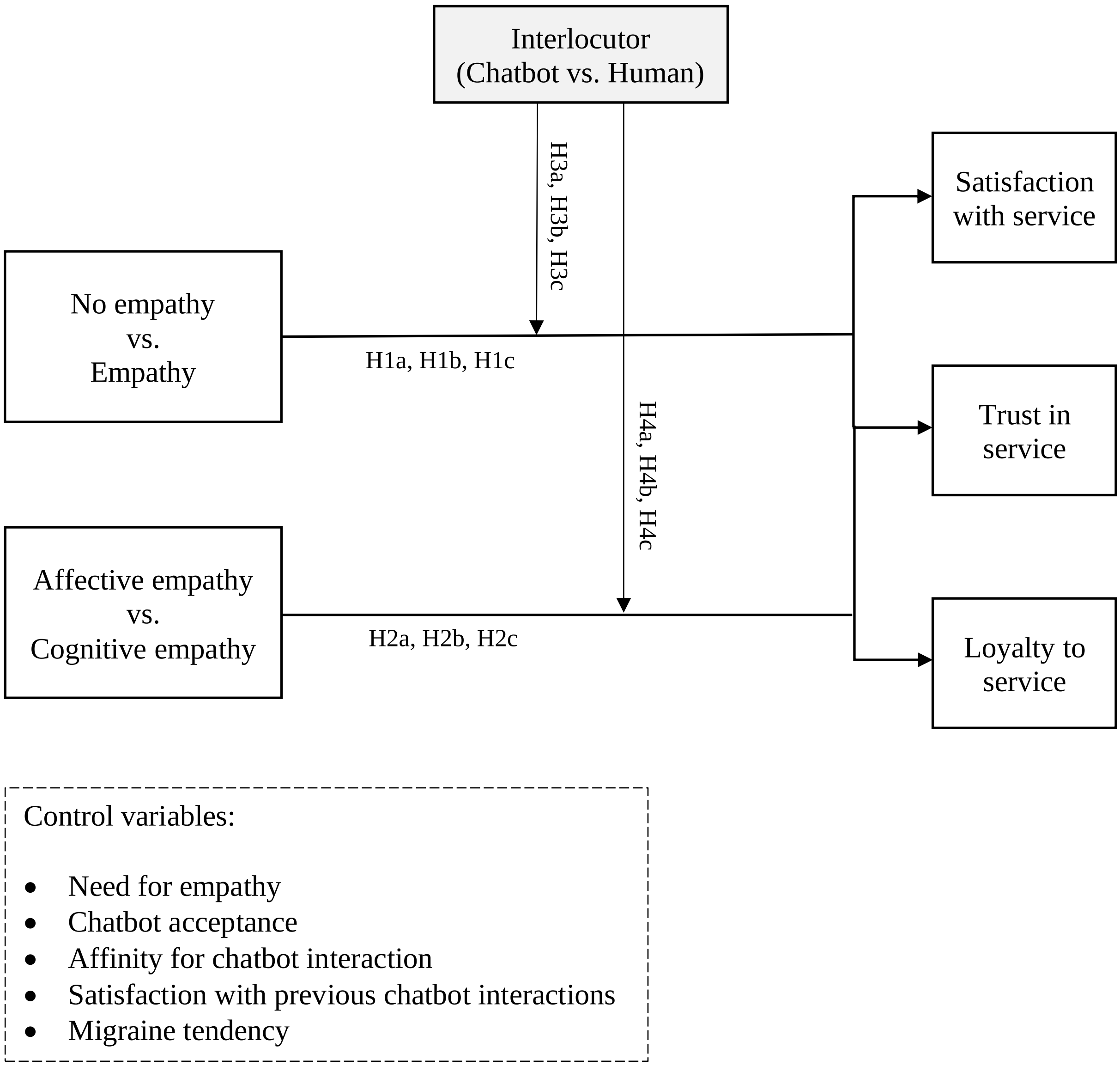
> ***H4b:*** *Having a chatbot as interlocutor (vs. a human) weakens the positive effect of affective empathy (vs. cognitive empathy) on the user's trust in the service.*

> ***H4c:*** *Having a chatbot as interlocutor (vs. a human) weakens the positive effect of affective empathy (vs. cognitive empathy) on the user's loyalty to the service.*

Figure 1 shows the resulting conceptual framework for this study.

Figure 1

The conceptual framework of this study



3. Methodology

3.1 Method

The conducted empirical research has an experimental design. This decision is based on two arguments (Bonoma, 1985; Yin, 1994). First, the research's aim comprises a "how" question, as it tries to understand the phenomenon of how empathy might affect the user's satisfaction with, trust in, and loyalty to the service of a THCB. Second, the phenomenon's complexity is moderate. The interaction with a THCB can fairly well be studied outside its natural setting, while variables like satisfaction, trust and loyalty are ideal for self-assessment. All in all, an experimental approach seems very suitable (Bonoma, 1985; Yin, 1994).

For the experiment, respondents are being randomly assigned to either a chatbot condition or a human condition. Each of the two conditions are further divided into the randomly assigned groups "*no empathy*", "*affective empathy*", and "*cognitive empathy*". Thus, the experiment follows a 2x3-between-subject-design. More specifically, the experiment starts by asking the respondents to imagine just having had a strong migraine attack and still feeling sick, which causes them to seek advice via a healthcare service app, recommended by their GP. After presenting this introductory scenario, respondents are being shown a simulated text-based conversation of them getting migraine-related advice from either a THCB (chatbot condition) or a healthcare professional (human condition). Once again, respondents are being asked to imagine having this conversation. During this dialog, the subgroups "no empathy" receive replies by the THCB/healthcare professional that solely focus on the health-related advice and information. In the subgroups "affective empathy", the THCB/healthcare professional additionally includes phrases such as "*Oh, I'm really sorry.*" or "*I totally feel your situation*." to portray the essence in affective empathy of feeling the user's state. The replies in the subgroups "cognitive empathy" instead comprise phrases such as "*Oh, I understand*." or "*I totally understand your situation*.", pointing out that the user's state is rather understood than felt. After the conversation, the participants are asked to answer various questions about the experienced interaction and themselves, resulting in the needed measurements.

The texts that are presented to the participants as the six manipulating conditions vary only slightly in length (chatbot conditions: 394 words, 453 words, 465 words; human conditions: 397 words, 456 words, 468 words). With a reading speed of 250 words per minute for an average reader (Fry, 1963), the biggest difference in number of words throughout the manipulations of 468 – 394 = 75 words is expected to demand an additional reading time of only 18 seconds. There is thus no need to integrate a waiting time for respondents with shorter manipulations for bias reduction, as a waiting time itself is a critical determinant of service satisfaction (Bielen, 2007) and would have a significant chance of causing the bigger bias.

Next, it is critical to make the presented situation and conversation as natural and realistic as possible, in order to improve the study's internal validity (Seltman, 2012). The choice of using migraine as the driving health concern for the interaction is supposed to contribute to that purpose. In Western Europe and the United States, roughly 11% of the population actually suffer from migraine (Goadsby, Lipton, & Ferrari, 2002). Although not proven by a lot of academic data, it can be further assumed that the accompanying symptoms such as a strong headache, nausea, vomiting or increased sensitivity to light, sound, or movement are relatable concerns most people have experienced at some point and thus know how to imagine. The health-related advice in the conversation is taken from popular websites and blogs that provide advice and relevant information regarding various health problems (Mayo Clinic, 2020; Orenstein, 2017; Schori, 2018).

Pretest. A pretest with 20 participants that took place from April 29th 2021 to April 30th 2021 showed that all conducted manipulations had a desired effect: First of all, respondents perceived both their chatbot or human conditions as such. Those who talked to a human scored an average of 5.17 (SD = 1.69) on a 7-point Likert scale that measured the perceived humaneness of the interlocutor. Respondents who talked to a chatbot only scored an average of 2.33 (SD = 1.66), clearly showing that the interlocutor was not seen as human.

Next, the manipulations for no empathy, affective empathy and cognitive empathy of the health adviser were effective. However, they were not statistically significant, which can most probably be explained by the low number of participants for the pretest. Still, respondents in the "No Empathy"-group perceived the least affective and cognitive empathy ($M_{\text{affective}}$ = 4.09; $SD_{\text{affective}}$ = 1.51; $M_{\text{cognitive}}$ = 4.19; $SD_{\text{cognitive}}$ = 1.31). Respondents in the "Affective Empathy"-group perceived more affective than cognitive empathy ($M_{\text{affective}}$ = 4.67; $SD_{\text{affective}}$ = 1.51; $M_{\text{cognitive}}$ = 4.42; $SD_{\text{cognitive}}$ = 1.24), while respondents in the "Cognitive Empathy"-group perceived more cognitive than affective empathy ($M_{\text{affective}}$ = 4.17; $SD_{\text{affective}}$ = 1.48; $M_{\text{cognitive}}$ = 5.00; $SD_{\text{cognitive}}$ = 1.38).

Lastly, respondents perceived the presented situation in the experiment as fairly realistic. A 7-point Likert scale measuring perceived realism indicated an average score of 4.31 (SD = 1.14). This seems sufficient for conducting further research when considering the still prevailing newness of the chatbot technology in healthcare.

3.2 Data collection and sample

The data for the main experiment was collected through a Qualtrics online survey, which was primarily distributed on social messengers and social media networks (Facebook, Instagram, WhatsApp), but also via mouth-to-mouth-marketing. The data collection took place from April 30th 2021 to May 5th 2021. As anybody can be affected by health issues and is thus a potential user of THCB's, the research did not focus on a specific group of people.

The extracted sample initially included 217 respondents. In order to increase the data's validity and ensure that participants took part with sufficient attention, respondents with extreme values for survey duration time were excluded. Therefore, the interquartile range (IQR) rule with a multiplier of 3 was used (Field, 2018). This means that values that fell below the 25 percentile (7 min; 26 sec) or exceeded the 75 percentile (12 min; 8 sec) by more than the threefold of the interquartile range of 4 minutes and 42 seconds (i.e., by 14 min; 6 sec) were deleted. As a result, all answers which exceeded the upper boundary of 26 minutes and 14 seconds were excluded. As this rule did not lead to a valid lower boundary, answers below the 5 percentile and above the 95 percentile are additionally being deleted. While this did not lead to further eliminations on the upper side, answers with a duration time below 4 minutes and 22 seconds were eliminated.

After this data cleansing, the final sample comprises 197 responses. With 133 women (67.5%) and 63 men (32.0%) participating, a surplus of female respondents can be identified. One participant (0.5%) choses the option "other/prefer not to say" when being asked about gender. The respondents' age ranges from 16 to 78 years. 49.8% and thus approximately half of the respondents are between 21 and 25 years old. The overall average age is 29.26 with a standard deviation of 11.75. Furthermore, as the research is mainly conducted in Germany, most respondents are German (85). Due to the far-reaching network effects of social media, the remaining nationalities are widely spread. The most represented countries of origin are the United States of America (19), the Netherlands (13), the United Kingdom (11), and India (6). Finally, the respondents' highest level of education is balanced with most people having obtained a Bachelor's degree (92), a Master's degree (33) or having completed high school/German "Abitur" (50).

3.3 Operationalization

Dependent variables. The measurement scale for the variable *satisfaction with service* is based on a 7-point semantic scale with nine items, used by Oliver and Swan (1989). It has been originally invented to measure a consumer's degree of satisfaction with another party with whom a transaction has occurred or relationship has developed. Some items are slightly reworded to match the settings of the conditions including the THCB, as well as the conditions including the human health adviser. Appendix 1 provides an overview of the original and altered items for this and all following scales.

For the variable *trust in service,* a 7-point Likert scale with four items by Chaudhuri and Holbrook (2001) is consulted. First created to measure a consumer's trust in a specific brand, it also undergoes a small adjustment in wording to provide a fitting measurement scale for the user’s trust in the experienced service.

The third and final dependent variable *loyalty to service* is measured with the help of a 7-point likelihood scale by Zeithaml, Berry, and Parasuraman (1996), comprising a total of five items. It initially consisted of two subdimensions in the form of *word-of-mouth communications* with three items and *repurchase intention* with two items. After adjusting the wording of the items, the subdimensions can be renamed to *word-of-mouth communications* and *intention to use service again* for the service setting of this study.

Control variables. In order to minimize bias caused by other factors during the research, it is controlled for the user’s need for empathy, satisfaction with previous chatbot interactions, affinity for chatbot interaction, and chatbot acceptance.

The user’s *need for empathy* is measured by using items from the need for emotional support subscale of the Interpersonal Orientation Scale (IOS; Hill, 1987). This subscale tends to correlate substantially with measures related to empathy and sociability, implying that it covers a dimension of desire for empathic behavior and feeling understood (Hill, 1987). The original 5-point Likert scale is adjusted to a 7-point Likert scale.

To measure the user’s *satisfaction with previous chatbot interactions*, the same scale by Oliver and Swan (1989) is consulted as for the variable *satisfaction with service.* Hence, this measurement also consists of a 7-point semantic scale with nine items.

The variable *affinity for chatbot interaction* is supposed to control for the user’s general affinity for and knowledge about chatbot interaction systems. Therefore, a 9-item scale by Franke, Attig and Wessel (2019) is consulted, which was originally developed to measure general affinity for technology interaction. To provide better comprehensibility for the respondent, only six out of the nine items are used, taking those that seem most suited for the setting. By slightly rewording the items, a chatbot setting is formed. Moreover, the 6-point Likert scale is extended to a 7-point Likert scale. This adjustment should not have any significant effects on the resulting data (Dawes, 2007; Leung, 2011). Item 5 is reversed.

Next, controlling for the user’s *chatbot acceptance* is done by using the fast form approach for technology acceptance measurements by Chin, Johnson and Schwarz (2008). This measurement is based on the Technology Acceptance Model by Davis (1989) and consists of a 12-item semantic differential scale. It states that technology acceptance is determined by perceived usefulness and perceived ease of use, and thus includes six items for each factor. The fast form approach provides customizable wording, making it possible to tailor it to the specific topic of chatbot acceptance. In order to minimize response fatigue among the participants, two items of each factor are omitted, choosing those with the least factor loadings and suitability for this specific setting. Thus, the final scale consists of eight items in total. Furthermore, it is slightly adjusted, turning the original 9-point scale into a 7-point scale.

Lastly, the user's *migraine tendency* is measured with a single question about how often the person experiences a migraine or strong headaches in real life. The measurement consists of a 7-point frequency scale reaching from 1 = "*never*" to 7 = "*always*".

Manipulation checks. The shown empathy by the advising interlocutor during the conversation acts as an independent variable that is getting manipulated. In order to measure that "no empathy", "affective empathy", and "cognitive empathy" are perceived as such and thus prove the effectiveness of the manipulation, an 8-item scale by Plank, Minton and Reid (1996) is borrowed. The scale was originally invented to measure perceived empathy in a sales setting. It covers both *perceived affective empathy* and *perceived cognitive empathy* by comprising four items for each dimension, making the scale perfectly appropriate for this study. Due to a high Cronbach alpha of .93 for this one-factor scale and thus good internal consistency, it seems suitable to be slightly reworded. The 5-point Likert scale is adjusted to a 7-point scale in order to contribute to a consistent measurement format throughout all variables. Once again, this adjustment should not have any significant effects on the resulting data (Dawes, 2007). Item 1 of the perceived affective empathy scale and item 3 of the perceived cognitive empathy scale are formulated negatively and are thus reversed.

The *interlocutor* itself functions as the potential moderating variable in the study's conceptual framework, and is getting manipulated by randomly assigning respondents to either a chatbot or a human being as a health adviser. In order to check this manipulation and show that respondents really perceive their interlocutor as either a programmed chatbot or a human being, a *perception of interlocutor* scale is created. This self-developed 7-point Likert scale consists of three items, e.g. "*My interlocutor is a real person.*". Thus, the items are formulated in a way that an effective chatbot condition should lead to low agreement, while an effective human condition should lead to high agreement.

Finally, the participants' *perceived realism* of the experimental setting is measured with the help of a 7-point Likert scale. Its four items are taken from a realism measurement scale by Busselle (2001). By slightly rewording the items, the original TV setting of the questions is adjusted to the healthcare service setting of this experiment. Items 2 and 4 are reversed.

3.4 Data analysis procedure

The extracted data is analyzed with the help of the statistics program SPSS and generally follows the statistics guides of Field (2018) and Hair et al. (2019). The actual analysis is preceded by a manipulation check. Here, it is tested whether the manipulations for the condition (Chatbot vs. Human), as well as for the shown empathy (no empathy vs. affective empathy vs. cognitive empathy) are perceived as such. Furthermore, the perceived realism of the experimental setup is checked.

Thereafter, the hypotheses for the main effects H1a, H1b, H1c, and H2a, H2b, H2c are tested. Hence, it is firstly tested whether empathy (vs. no empathy) shown by the interlocutor increases the user's satisfaction with, trust in, and loyalty to the service. Next, it is checked whether affective (vs. cognitive) empathy increases the user's satisfaction, trust, and loyalty. During the analyses, it shall be controlled for the user's need for empathy, chatbot acceptance, affinity for chatbot interactions, satisfaction with previous chatbot interactions, and migraine tendency. The hypotheses are tested with the help of several one-way ANOVAs and ANCOVAs. A one-way ANOVA can compare the means between two groups, while a one-way ANCOVA additionally controls for covariates (control variables) (Field, 2018). It is to mention that only 125 out of the total 197 respondents had previous chatbot experiences and could thus give answers for the five control variables. This causes further differences in sample size across the groups and analyses.

Finally, potential moderating effects of the interlocutor on the main effects are tested. In particular, it is checked whether the effects of empathy on the user's satisfaction, trust, and loyalty are moderated by the user's interlocutor being either a chatbot or a human. Therefore, the analysis again starts with conducting one-way ANOVAs and one-way ANCOVAs in order to test main effects. This time, however, these are conducted separately for the chatbot and human condition. This shall allow for a first overview of potential differences and thus moderating effect of the user's interlocutor. Afterwards, two-way ANOVAs and two-way ANCOVAs are conducted to test for statistical significance. A two-way ANOVA can determine the interaction effect between two independent variables on a dependent variable, while a two-way ANCOVA can additionally control for covariates (Field, 2018). Once again, it is controlled for the user's need for empathy, chatbot acceptance, affinity for chatbot interactions, satisfaction with previous chatbot interactions, and migraine tendency.

3.5 Research ethics

Throughout the whole process of this research, high ethical standards were met. The ethical framework for protecting the rights of human subjects in fieldwork can be summarized into three main areas (Alsmadi, 2008): The right to be informed, the right to privacy and confidentiality, and deception and harm.

First, in order to protect participants' right to be informed, the conducted online questionnaire clearly stated the purpose of the research at the beginning, letting respondents then decide whether a participation is desired. Furthermore, sufficient information about the process of the research and usage of the collected date was provided. Second, the right to privacy and confidentiality was ensured by conducting the research on an anonymous basis, only using the data for intended and disclosed purpose. Lastly, any deception and harm of the respondents or their provided data was meant to be avoided by only giving truthful information about research and researcher, and by designing and conducting the survey in a way that by no means aimed to cause any physical or psychological harm to any respondent.

4. Results

4.1 Manipulation checks

The manipulation check leads to the following results: Firstly, it cannot be fully confirmed that the respondents perceived their chatbot or human conditions as such. Those who talked to a human scored an average of 4.03 (SD = 1.69) on a 7-point Likert scale that measured the perceived humaneness of the interlocutor. Respondents who talked to a chatbot scored a slightly lower average of 3.38 (SD = 1.54). However, as the two conditions were quite clearly communicated in the experiment, it is possible that the respondents actually understood their conditions and only scored similar results due to two reasons: The used scale that measured perceived humaneness of the interlocutor is self-developed and thus has no proof of validity or reliability. Moreover, the simulated health professional that was supposed to act as a human interlocutor could still be seen as not human since it is not a real person, ultimately leading to misinterpretations of the measurement scale.

Second, the effectiveness of the manipulations for no empathy, affective empathy and cognitive empathy of the health adviser cannot be fully confirmed after conducting a one-way ANOVA (Field, 2018). As desired, respondents in the "Affective Empathy"-group had the highest scores for perceived affective empathy ($M_{\text{affective}}$ = 4.49; $SD_{\text{affective}}$ = 1.31) compared to the groups "No Empathy" ($M_{\text{affective}}$ = 4.37; $SD_{\text{affective}}$ = 1.14) and "Cognitive Empathy"

($M_{affective}$ = 3.95; $SD_{affective}$ = 1.31). However, the "Affective Empathy"-group also had the highest scores for perceived cognitive empathy ("Affective Empathy"-group: $M_{cognitive}$ = 4.72; $SD_{cognitive}$ = 1.31; "No Empathy"-group: $M_{cognitive}$ = 4.17; $SD_{cognitive}$ = 1.26; "Cognitive Empathy"-group: $M_{cognitive}$ = 4.23; $SD_{cognitive}$ = 1.60), which, moreover, was even higher than the group's score for perceived affective empathy. Furthermore, it can be noted that the differences between groups are just statistically significant on a significance level of α = .05 for both affective empathy (p = .043) and cognitive empathy (p = .049) (Field, 2018). The missing desired outcome regarding specific empathy manipulations should not be viewed too critically. It is possible that respondents could not truly evaluate their actual perception when being asked to retrospectively reflect on the experience. In fact, although manipulation checks can be helpful, they do not always give valid and reliable results. A manipulation check is not just a measure but many times an additional event for participants, potentially influencing their psychological thought process and thus affecting their self-evaluations during the measurement (Hauser, Ellsworth, & Gonzalez, 2018).

Lastly, the manipulation check showed that respondents perceived the presented situation in the experiment as fairly realistic. A 7-point Likert scale measuring perceived realism indicated an average score of 4.26 (SD = 1.36) throughout the whole sample. Just like in the pretest, this seems sufficient for conducting further research as it appears that the still prevailing newness of the chatbot technology in healthcare could be the reason for moderate perceptions of realism for the presented scenario. However, respondents who had a chatbots as interlocutor even perceived the situation as slightly more realistic than those who talked to a human (Chatbot group: $M_{perceived\_realism}$ = 4.45; $SD_{perceived\_realism}$ = 1.35; Human group: $M_{perceived\_realism}$ = 4.07; $SD_{perceived\_realism}$ = 1.35). It seems like it is rather talking to a healthcare professional via an app which lacks realism for some respondents, while talking to a chatbot in this app context is perceived as more realistic.

4.2 Analysis

4.2.1 Assumptions

In order to analyze data with a one-way and two-way ANOVA and ANCOVA, several assumptions have to be met, which is the case for the used data. First of all, the dependent variables and covariate variables are measured on a continuous scale (7-point scale). Furthermore, the included independent variables consist of two categorical, independent groups with observations being independent as no participant is in more than one group. Creating boxplots in SPSS for each dependent and covariate variable shows that there are no outliers except for two extreme values for the dependent variable *satisfaction*. However, these can be retained as there is no proof that these values are not representative for any observations in the population (Hair et al., 2019).

A Shapiro-Wilk test for each dependent and covariate variable indicates significance for some variables, implying that a normal distribution for each category of the independent variables cannot be assumed. Nonetheless, a violation of the normality assumption does not have a big impact on larger sample sizes (Hair et al., 2019). Moreover, a Levene's test can confirm homogeneity of variances for almost all variables. Only the variances of the variables *satisfaction* and *need for empathy* do not seem to be homogeneous as they exceed the critical p-value of .05 (Satisfaction: $p = .021$; Need for empathy: $p = .030$). Next, scatterplots indicate that the covariates are linearly related to the dependent variables at each level of the independent variables and that homoscedasticity is given. Finally, homogeneity of regression slopes can be confirmed for almost every covariate and independent variable. The only exception is the covariate *Chatbot acceptance*, which constantly shows significant p-values of under .05 for the test of between-subject effects. Overall, only some assumptions are mildly violated. Nevertheless, it still seems appropriate to conduct one-way and two-way ANOVAs and ANCOVAs to test the hypotheses (Hair et al., 2019)

4.2.2 Hypotheses tests for the main effects

In this subchapter, the hypotheses for the main effects H1a, H1b, H1c, and H2a, H2b, H2c are tested. Thus, it is firstly tested whether empathy (vs. no empathy) from the interlocutor increases the user's satisfaction with, trust in, and loyalty to the service. Afterwards, the same is done for affective vs cognitive empathy. During the analyses, it is controlled for the user's need for empathy, chatbot acceptance, affinity for chatbot interactions, satisfaction with previous chatbot interactions, and migraine tendency.

No empathy vs. Empathy. The first one-way ANOVA and one-way ANCOVA compare the means for the user's satisfaction, trust, and loyalty between the "No empathy"-group and the "Empathy"-group. Table 1 sums up the most relevant results of these analyses.

First looking into satisfaction, a slight positive effect of empathy can be identified. However, neither the one-way ANOVA, nor the one-way ANCOVA show statistical significance for the effect, as the p-values exceed the critical value of .05 (Field, 2018). The assumption that empathy (vs. no empathy) leads to more satisfaction cannot be confirmed. H1a is thus not supported.

Contrary to the expectations, the user's trust in the service seemed to be negatively affected by empathy. Moreover, the one-way ANCOVA indicates a marginally higher mean difference compared to the one-way ANOVA. Nevertheless, the mean differences are still very small, leading to both analyses showing no statistical significance for the effect (Field, 2018). Empathy (vs. no empathy) does not seem to increase the user's trust in the service. Therefore, H1b is not supported.

Further, no effect of empathy on the user's loyalty to the service can be identified. However, when specifically looking into the subdimensions of loyalty, it can be seen that the user's intention to use the service again was even slightly decreased by empathy, while the user's word-of-mouth communications did not seem to be affected. As neither of the effects have statistical significance (Field, 2018), the assumption that empathy (vs. no empathy) increases the user's loyalty to the service cannot be confirmed. H1c is not supported.

Table 1

Main effects – "No empathy" vs. "Empathy" on Satisfaction, Trust, and Loyalty

	No empathy – Mean (Std. Dev.)	Empathy – Mean (Std. Dev.)	ANOVA/ANCOVA
Satisfaction			
ANOVA (N=197)	5.15 (1.44)	5.19 (1.30)	$F(1, 195) = .051, p = .822$
ANCOVA (N=125)	5.03 (1.37)	5.21 (1.22)	$F(1, 118) = .045, p = .833$
Trust			
ANOVA (N=197)	4.68 (1.34)	4.56 (1.36)	$F(1, 195) = .334, p = .564$
ANCOVA (N=125)	4.66 (1.36)	4.46 (1.37)	$F(1, 118) = 1.126, p = .291$
Loyalty			
ANOVA (N=197)	4.79 (1.63)	4.69 (1.65)	$F(1, 195) = .153, p = .697$
ANCOVA (N=125)	4.69 (1.66)	4.63 (1.63)	$F(1, 118) = .205, p = .652$
Loyalty_WOM			
ANOVA (N=197)	5.00 (1.68)	5.00 (1.62)	$F(1, 195) = .000, p = .991$
ANCOVA (N=125)	4.83 (1.70)	4.94 (1.58)	$F(1, 118) = .040, p = .842$
Loyalty_Intention			
ANOVA (N=197)	4.47 (1.80)	4.23 (1.81)	$F(1, 195) = .815, p = .368$
ANCOVA (N=125)	4.48 (1.80)	4.16 (1.80)	$F(1, 118) = 1.706, p = .194$

Affective empathy vs. Cognitive Empathy. Next, the effects of affective empathy compared to cognitive empathy on the user's satisfaction, trust, and loyalty are tested. Table 2 shows the most relevant results from these main effect tests.

As expected, affective empathy seemed to increase the user's satisfaction with the service compared to cognitive empathy. The conducted one-way ANOVA even indicates statistical significance for the effect with a p-value lower than .05 ($F(1, 124) = 5.110$, $p = .026$) (Field, 2018). When additionally controlling for the covariates with a one-way ANCOVA, the mean difference shrinks, ultimately taking away statistical significance ($F(1, 74) = .221$, $p = .639$) (Field, 2018). The reason for this could lie in the user's chatbot acceptance, as it is the only covariate that appeared to significantly affect the user's satisfaction ($F(1, 74) = 7.506$, $p = .008$) (Field, 2018). Furthermore, it has to be mentioned that the one-way ANCOVA (N=81) included a smaller sample size than the one-way ANOVA (N=126), giving the analysis an overall lower statistical power (Field, 2018). Still, it cannot be assumed that affective empathy leads to more satisfaction than cognitive empathy does. H2a is thus not supported.

The results are similar for the user's trust in the service. Affective empathy generally led to higher trust scores than cognitive empathy did. While the one-way ANOVA indicates statistical significance for the effect ($F(1, 124) = 5.002$, $p = .027$), the one-way ANCOVA does not ($F(1, 74) = 1.729$, $p = .193$) (Field, 2018). Once again, the user's chatbot acceptance seemed to be the decisive influence, showing a significant effect on trust ($F(1, 74) = 7.098$, $p = .009$) (Field, 2018). Although controlling for this covariate still leads to a noticeably higher mean in trust for the "Affective empathy"-group compared to the "Cognitive empathy"-group, the missing statistical significance in the one-way ANCOVA should not be neglected (Field, 2018). Therefore, H2b is not supported.

Finally, the user's loyalty also experienced higher scores in the "Affective empathy"-group than in the "Cognitive empathy"-group. The same applied to the two subdimensions of loyalty, word-of-mouth communications and intention to use the service again. This time, however, both the conducted one-way ANOVA and one-way ANCOVA indicate no statistical significance (Field, 2018). Looking into the one-way ANCOVA, the user's chatbot acceptance again showed a significant effect on overall loyalty ($F(1, 74) = 7.792$, $p = .007$) (Field, 2018). In conclusion, it cannot be assumed that affective empathy increases the user's loyalty compared to cognitive empathy. Thus, H2c is not supported.

Table 2

Main effects – "Affective empathy" vs. "Cognitive empathy" on Satisfaction, Trust, and Loyalty

	Affective empathy – Mean (Std. Dev.)	Cognitive empathy – Mean (Std. Dev.)	ANOVA/ANCOVA
Satisfaction			
ANOVA (N=126)	5.44 (1.08)	4.92 (1.47)	$F(1, 124) = 5.110$, $\boldsymbol{p = .026}$
ANCOVA (N=81)	5.34 (1.04)	5.08 (1.36)	$F(1, 74) = .221$, $p = .639$
Trust			
ANOVA (N=126)	4.82 (1.32)	4.28 (1.36)	$F(1, 124) = 5.002$, $\boldsymbol{p = .027}$
ANCOVA (N=81)	4.68 (1.44)	4.26 (1.29)	$F(1, 74) = 1.729$, $p = .193$
Loyalty			
ANOVA (N=126)	4.86 (1.69)	4.51 (1.60)	$F(1, 124) = 1.429$, $p = .234$
ANCOVA (N=81)	4.69 (1.79)	4.57 (1.48)	$F(1, 74) = .002$, $p = .962$
Loyalty_WOM			
ANOVA (N=126)	5.13 (1.67)	4.86 (1.56)	$F(1, 124) = .873$, $p = .352$
ANCOVA (N=81)	4.91 (1.74)	4.97 (1.44)	$F(1, 74) = .292$, $p = .591$
Loyalty_Intention			
ANOVA (N=126)	4.45 (1.82)	3.98 (1.77)	$F(1, 124) = 2.156$, $p = .145$
ANCOVA (N=81)	4.37 (1.95)	3.96 (1.63)	$F(1, 74) = .672$, $p = .415$

4.2.3 Hypotheses tests for moderating effect

Now, potential moderating effects of the interlocutor are tested. Therefore, it is checked whether the effects of empathy on the user's satisfaction, trust, and loyalty are moderated by the user's interlocutor being either a chatbot or a human. Once again, it is controlled for the user's need for empathy, chatbot acceptance, affinity for chatbot interactions, satisfaction with previous chatbot interactions, and migraine tendency.

No empathy vs. Empathy. First, the focus is on no empathy vs. empathy. Table 3 shows the most relevant results of the main effect tests in the chatbot group, while table 4 shows the same for the human group. Table 5 presents the results of the actual interaction effect tests.

Only looking at respondents who talked to a chatbot, empathy slightly increased the user's satisfaction compared to no empathy. Nonetheless, neither the one-way ANOVA, nor the one-way ANCOVA show statistical significance for the effect (Table 3) (Field, 2018). In contrast, respondents who talked to a human rather experienced a minor decrease in satisfaction when being advised with empathy. Again, the effect still misses statistical significance in both the one-way ANOVA and one-way ANCOVA (Table 4) (Field, 2018). The predicted effect that a chatbot (vs. a human) amplifies a positive effect of empathy on satisfaction can thus be partially recognized. However, after conducting a two-way ANOVA and ANCOVA, no statistical significance is found (Field, 2018), i.e., the independent variable of no empathy vs. empathy does not show an interaction effect with the interlocutor with regards to satisfaction (Table 5). Overall, H3a is not supported.

The user's trust seemed to be slightly decreased by empathy when the interlocutor was a chatbot, but not significantly (Table 3) (Field, 2018). The same was recognizable for the human condition (Table 4). Hence, relying on the analyses of the main effects, no moderating effect of the interlocutor can be assumed. This is confirmed by both the two-way ANOVA and ANCOVA, which indicate no significant interaction effect (Table 5) (Field, 2018). The assumption that a chatbot (vs. a human) amplifies a positive effect of empathy on trust can thus not be confirmed. H3b is not supported.

Lastly, the results do not change majorly when taking a look at the user's loyalty. Generally, empathy appeared to lead to a small decrease in loyalty for both the chatbot (Table 3) and human condition (Table 4) (Field, 2018). Only the one-way ANCOVA for the human condition can detect a marginal increase in loyalty when the respondent is advised with empathy. More striking are the results for the subdimension of the user's intention to use the service again. Here, a chatbot with empathy decreased the user's reuse intention quite noticeably (Table 3). A human with empathy, however, did not lower the reuse intention (Table 4). Still, no interaction effect can be identified (Table 5) (Field, 2018). The hypothesis that a chatbot (vs. a human) amplifies a positive effect of empathy on loyalty can thus not be confirmed. Hence, H3c is not supported.

Table 3

Main effects – "No empathy" vs. "Empathy" on Satisfaction, Trust, and Loyalty (only ***chatbot*** *condition)*

	No empathy – Mean (Std. Dev.)	Empathy – Mean (Std. Dev.)	ANOVA/ANCOVA
Satisfaction			
ANOVA (N=98)	5.10 (1.65)	5.37 (1.22)	F(1, 96) = .830, p = .364
ANCOVA (N=59)	5.12 (1.72)	5.37 (1.08)	F(1, 52) = .182, p = .672
Trust			
ANOVA (N=98)	4.64 (1.45)	4.70 (1.26)	F(1, 96) = .050, p = .823
ANCOVA (N=59)	4.88 (1.42)	4.58 (1.22)	F(1, 52) = .335, p = .565
Loyalty			
ANOVA (N=98)	4.90 (1.66)	4.75 (1.60)	F(1, 96) = .208, p = .649
ANCOVA (N=59)	5.12 (1.76)	4.67 (1.57)	F(1, 52) = .650, p = .424
Loyalty_WOM			
ANOVA (N=98)	5.08 (1.67)	5.07 (1.57)	F(1, 96) = .000, p = .999
ANCOVA (N=59)	5.16 (1.74)	4.99 (1.53)	F(1, 52) = .058, p = .811
Loyalty_Intention			
ANOVA (N=98)	4.66 (1.81)	4.26 (1.77)	F(1, 96) = 1.063, p = .305
ANCOVA (N=59)	5.06 (1.88)	4.19 (1.75)	F(1, 52) = 2.380, p = .129

Table 4

Main effects – "No empathy" vs. "Empathy" on Satisfaction, Trust, and Loyalty (only ***human*** *condition)*

	No empathy – Mean (Std. Dev.)	Empathy – Mean (Std. Dev.)	ANOVA/ANCOVA
Satisfaction			
ANOVA (N=99)	5.18 (1.26)	4.98 (1.37)	$F(1, 97) = .518$, $p = .473$
ANCOVA (N=66)	4.97 (1.13)	5.03 (1.34)	$F(1, 59) = .021$, $p = .885$
Trust			
ANOVA (N=99)	4.71 (1.27)	4.40 (1.46)	$F(1, 97) = 1.161$, $p = .284$
ANCOVA (N=66)	4.53 (1.33)	4.33 (1.53)	$F(1, 59) = .509$, $p = .478$
Loyalty			
ANOVA (N=99)	4.70 (1.62)	4.63 (1.71)	$F(1, 97) = .035$, $p = .851$
ANCOVA (N=66)	4.42 (1.58)	4.58 (1.70)	$F(1, 59) = .033$, $p = .856$
Loyalty_WOM			
ANOVA (N=99)	4.94 (1.71)	4.92 (1.68)	$F(1, 97) = .004$, $p = .952$
ANCOVA (N=66)	4.63 (1.67)	4.88 (1.65)	$F(1, 59) = .157$, $p = .694$
Loyalty_Intention			
ANOVA (N=99)	4.33 (1.79)	4.19 (1.86)	$F(1, 97) = .120$, $p = .730$
ANCOVA (N=66)	4.11 (1.68)	4.13 (1.87)	$F(1, 59) = .021$, $p = .885$

Table 5

Interaction effects between "No empathy vs. Empathy" and "Chatbot vs. Human" on Satisfaction, Trust, and Loyalty

	ANOVA (N=197)	ANCOVA (N=125)
Satisfaction	$F(1, 193) = 1.344$, $p = .248$	$F(1, 116) = .002$, $p = .965$
Trust	$F(1, 193) = .837$, $p = .361$	$F(1, 116) = .028$, $p = .868$
Loyalty	$F(1, 193) = .038$, $p = .846$	$F(1, 116) = 1.062$, $p = .305$
Loyalty_WOM	$F(1, 193) = .002$, $p = .967$	$F(1, 116) = .449$, $p = .504$
Loyalty_Intention	$F(1, 193) = .250$, $p = .618$	$F(1, 116) = 2.051$, $p = .155$

Affective empathy vs. Cognitive empathy. Lastly, the focus is on affective empathy vs. cognitive empathy. Table 6 shows the most relevant results of the main effect tests in the chatbot group, while table 7 shows the same for the human group. Table 8 presents the results of the actual interaction effect tests.

While affective and cognitive empathy did not appear to significantly influence the user's satisfaction in the chatbot condition (Table 6), the results deviate greatly in the human condition. Here, both the one-way ANOVA and ANCOVA show that affective empathy led to considerably more satisfaction than cognitive empathy did (Table 7). The one-way ANOVA even indicates statistical significance ($F(1, 57) = 5.798$, $p = .019$) (Field, 2018). Still, the two-way ANOVA and ANCOVA cannot find a significant interaction effect between the type of empathy and the interlocutor with regards to satisfaction (Table 8) (Field, 2018). Therefore, the assumption that having a chatbot (vs. a human) as interlocutor weakens the positive effect of affective empathy (vs. cognitive empathy) on satisfaction encounters supporting tendencies in the results. However, due to missing statistical significance, H4a is not supported.

The results look very similar for the user's trust. Affective empathy and cognitive empathy did not have very differing effects on trust when the interlocutor is a chatbot (Table 6). In contrast, for the human condition, affective empathy led to substantially more trust than cognitive empathy did (Table 7). The one-way ANOVA again indicates a significant main effect ($F(1, 57) = 6.234$, $p = .015$) (Field, 2018). Nevertheless, no statistical significance is found after analyzing the interaction effect with a two-way ANOVA and ANCOVA (Table 8) (Field, 2018). Just like for satisfaction, the assumption for trust finds supporting tendencies. Still, H4b cannot be supported.

Finally, the largest differences between the chatbot and human condition can be identified in the effects of affective vs. cognitive empathy on the user's loyalty to the service. When respondents talked to a chatbot, cognitive empathy led to significantly more loyalty than affective empathy. This especially applied to the user's word-of-mouth communications, but also to the user's intention to use the service again. The opposite was the case for the human condition. Here, affective empathy led to drastically more loyalty, which is supported by the one-way ANOVA indicating high statistical significance ($F(1, 57) = 5.689$, $p = .020$) (Field, 2018). A significant moderating effect of the interlocutor thus seems to be present. This can be confirmed by the conducted two-way ANOVAs and ANCOVAs. The ANOVAs show statistical significance for the interaction effects on loyalty ($F(1, 122) = 4.734$, $p = .032$), as well as on both subdimensions of loyalty (Loyalty_WOM: $F(1, 122) = 4.790$, $p = .031$; Loyalty_Intention: $F(1, 122) = 4.032$, $p = .047$) (Field, 2018). The ANCOVAs only indicate significance for word-of-mouth communications ($F(1, 72) = 4.445$, $p = .038$). However, the p-values for reuse intention and especially loyalty overall just barely exceed the critical threshold of .05 (Loyalty: $F(1, 72) = 3.647$, $p = .060$; Loyalty_Intention: $F(1, 72) = 2.350$, $p = .130$). The missing statistical significance compared to the ANOVAs could very well be explained by the smaller sample size (Field, 2018). Hence, it can be assumed that having a chatbot as interlocutor (vs. a human) weakens the positive effect of affective empathy (vs. cognitive empathy) on the user's loyalty to the service. Thus, H4c is supported.

Table 6

Main effects – "Affective empathy" vs. "Cognitive empathy" on Satisfaction, Trust, and Loyalty (only ***chatbot*** *condition)*

	Affective empathy – Mean (Std. Dev.)	Cognitive empathy – Mean (Std. Dev.)	ANOVA/ANCOVA
Satisfaction			
ANOVA (N=67)	5.49 (1.04)	5.25 (1.40)	$F(1, 65) = .676$, $p = .414$
ANCOVA (N=42)	5.23 (1.04)	5.49 (1.12)	$F(1, 35) = .012$, $p = .914$
Trust			
ANOVA (N=67)	4.80 (1.21)	4.59 (1.33)	$F(1, 65) = .441$, $p = .509$
ANCOVA (N=42)	4.51 (1.34)	4.63 (1.13)	$F(1, 35) = .116$, $p = .736$
Loyalty			
ANOVA (N=67)	4.63 (1.74)	4.88 (1.45)	$F(1, 65) = .375$, $p = .542$
ANCOVA (N=42)	4.27 (1.86)	5.00 (1.23)	$F(1, 35) = 1.284$, $p = .265$
Loyalty_WOM			
ANOVA (N=67)	4.92 (1.75)	5.24 (1.37)	$F(1, 65) = .670$, $p = .416$
ANCOVA (N=42)	4.47 (1.83)	5.42 (1.10)	$F(1, 35) = 2.880$, $p = .099$
Loyalty_Intention			
ANOVA (N=67)	4.20 (1.85)	4.33 (1.71)	$F(1, 65) = .086$, $p = .770$
ANCOVA (N=42)	3.97 (1.99)	4.37 (1.54)	$F(1, 35) = .119$, $p = .732$

Table 7

Main effects – "Affective empathy" vs. "Cognitive empathy" on Satisfaction, Trust, and Loyalty (only ***human*** *condition)*

	Affective empathy – Mean (Std. Dev.)	Cognitive empathy – Mean (Std. Dev.)	ANOVA/ANCOVA
Satisfaction			
ANOVA (N=59)	5.38 (1.13)	4.55 (1.49)	$F(1, 57) = 5.798$, $\mathbf{p = .019}$
ANCOVA (N=39)	5.44 (1.07)	4.58 (1.48)	$F(1, 32) = .993$, $p = .326$
Trust			
ANOVA (N=59)	4.83 (1.45)	3.93 (1.34)	$F(1, 57) = 6.234$, $\mathbf{p = .015}$
ANCOVA (N=39)	4.84 (1.54)	3.80 (1.35)	$F(1, 32) = 1.537$, $p = .224$
Loyalty			
ANOVA (N=59)	5.12 (1.62)	4.09 (1.67)	$F(1, 57) = 5.689$, $\mathbf{p = .020}$
ANCOVA (N=39)	5.09 (1.67)	4.04 (1.60)	$F(1, 32) = 1.445$, $p = .238$
Loyalty_WOM			
ANOVA (N=59)	5.37 (1.58)	4.43 (1.68)	$F(1, 57) = 4.871$, $\mathbf{p = .031}$
ANCOVA (N=39)	5.32 (1.59)	4.42 (1.63)	$F(1, 32) = .910$, $p = .347$
Loyalty_Intention			
ANOVA (N=59)	4.74 (1.78)	3.59 (1.79)	$F(1, 57) = 6.140$, $\mathbf{p = .016}$
ANCOVA (N=39)	4.75 (1.90)	3.47 (1.65)	$F(1, 32) = 2.127$, $p = .154$

Table 8

Interaction effects between "Affective empathy vs. Cognitive empathy" and "Chatbot vs. Human" on Satisfaction, Trust, and Loyalty

	ANOVA (N=126)	ANCOVA (N=81)
Satisfaction	$F(1, 122) = 1.640$, $p = .203$	$F(1, 72) = 1.492$, $p = .226$
Trust	$F(1, 122) = 2.187$, $p = .142$	$F(1, 72) = 1.517$, $p = .222$
Loyalty	$F(1, 122) = 4.734$, $\mathbf{p = .032}$	$F(1, 72) = 3.647$, $p = .060$
Loyalty_WOM	$F(1, 122) = 4.790$, $\mathbf{p = .031}$	$F(1, 72) = 4.445$, $\mathbf{p = .038}$
Loyalty_Intention	$F(1, 122) = 4.032$, $\mathbf{p = .047}$	$F(1, 72) = 2.350$, $p = .130$

4.3 Post-hoc analysis

After conducting the main analyses, an additional post-hoc can lead to an even better understanding of the collected data.

Firstly, it is worth reviewing the role of the control variables again. Throughout all analyses, no control variable shows a significant main effect on the dependent variables. The only exception is the user's chatbot acceptance. Some of these significant main effects have already been briefly mentioned. However, a detailed and complete review has not been provided, which is now done in the following. When looking at the main effects of no empathy vs. empathy only for those respondents who talked to a chatbot, the user's chatbot acceptance shows a significant main effect on the user's trust ($F(1, 52) = 11.900$, $p = .001$) and loyalty ($F(1, 52) = 6.402$, $p = .014$), however, not satisfaction ($F(1, 52) = .433$, $p = .514$) (Field, 2018). Looking at the main effects of affective vs. cognitive empathy only for those respondents who talked to a chatbot, a significant main effect of the user's chatbot acceptance can only be identified for the user's trust ($F(1, 35) = 6.996$, $p = .012$). It can be concluded that especially the user's trust in the chatbot service seems to be majorly affected by the user's general chatbot acceptance. Interestingly, when reviewing the same for those respondents who talked to a human healthcare professional, the user's chatbot acceptance only has a significant main effect on the user's satisfaction with the service. This applies to both the no empathy vs. empathy ($F(1, 59) = 7.490$, $p = .008$) analysis, as well as the affective vs. cognitive empathy analysis ($F(1, 32) = 4.242$, $p = .048$).

Second, this post-hoc analysis gives room to closer examine a potential causal interrelation of the dependent variables satisfaction, trust, and loyalty, as it has been assumed in the theoretical foundation of this thesis. Most studies indicate that satisfaction leads to loyalty, while trust acts as a mediator (Kassim & Abdullah, 2010; Morgan & Hunt, 1994). Hence, this potential relationship is tested for the present data. This is done with the help of the PROCESS extension program for SPSS. PROCESS is a logistic regression path analysis modeling tool and can estimate direct and indirect effects in a mediator model (Hayes, 2018). In this case, it estimates the direct effect of satisfaction on loyalty, as well as the indirect effect of satisfaction on loyalty via trust. Moreover, this is done for each the chatbot (N = 98) and the human (N = 99) condition.

First looking at the chatbot condition, a significant total effect of the user's satisfaction on the user's loyalty can be detected, i.e., satisfaction is a significant predictor of loyalty. The estimated increase in the user's loyalty score is .8010 per unit of satisfaction ($\beta = .8010$; $t(96) = 9.0708$; $p < .001$). Satisfaction likewise explains a significant portion of the variance in loyalty

($R^2 = .4615$; $F(1, 96) = 82.279$, $p < .001$). More particularly, 45.5% of this relationship between satisfaction and loyalty is explained by a significant direct effect (p = .001), while 54.5% are explained by a significant indirect effect via trust. The significance of this mediating effect of trust between satisfaction and loyalty is proven by the bootstrap intervals not including zero (Hayes, 2018).

Conducting the same regression analysis for the human condition, the results are similar. Satisfaction is again a significant predictor of loyalty. The estimated increase in the user's loyalty score is 1.0614 per unit of satisfaction ($\beta = 1.0614$; $t(97) = 15.4416$; $p < .001$). Satisfaction further explains a significant portion of the variance in loyalty ($R^2 = .7108$; $F(1, 97) = 238.4417$, $p < .001$). This time, however, 60.64% of this relationship between satisfaction and loyalty is explained by a significant direct effect (p < .001), while only 39.36% are explained by a significant indirect effect via trust. Once again, the significance of this mediating effect of trust between satisfaction and loyalty is proven by the bootstrap intervals not including zero (Hayes, 2018).

The existence of a causal relationship between satisfaction and loyalty, as well as a mediating role of trust can thus be confirmed. Furthermore, the mediating effect of the user's trust seems to be stronger for those respondents who talked to a chatbot.

4.4 Summary

Despite only one out of the twelve hypotheses being statistically confirmed, the results of the analyses give supporting evidence for many of the formulated assumptions, and further provide thought-provoking insights into the field of empathy in healthcare service and chatbots.

Regardless of the user's interlocutor, it was predicted that empathy, compared to no empathy, would generally increase the user's satisfaction with, trust in, and loyalty to the healthcare service. Yet this was only the case for satisfaction (H1a), while trust (H1b) and loyalty (H1c) even slightly suffered. It should not be neglected that these first observations missed any statistical significance, which is why they should be viewed with caution. As empathy has already been found to rather positively influence satisfaction, but also trust and loyalty (Homburg, Wieseke, & Bornemann, 2009; Feng, Lazar, & Preece, 2004; Bahadur et al., 2019), it could be assumed that the experiment was just missing a representative enough sample to confirm that. However, a more reasonable explanation could be that for this part of the experiment neither the different dimensions of empathy, nor the different interlocutors of the user have been taken into consideration. Hence, the missing specification might have led to the assumptions not being greatly supported.

When looking at the different effects of affective empathy and cognitive empathy, while still disregarding the interlocutor as an influencing factor, the results were in align with what has been predicted. Despite missing statistical significance, affective empathy visibly led to more satisfaction (H2a), trust (H2b), and loyalty (H2c) than cognitive empathy. The perception of someone else really feeling the own inner state seems to be more appreciated by the user than the sole perception of being cognitively understood. Noteworthy for the results of this part are the strong main effects of the user's chatbot acceptance on satisfaction, trust, and loyalty. The other covariates in the form of the user's need for empathy, affinity for chatbot interactions, satisfaction with previous chatbot interactions, and migraine tendency did not show any main effects.

The results became clearer after it has additionally been distinguished between the user's interlocutor being either a chatbot or a human. It was noticeable that the earlier detected positive effect of empathy (vs. no empathy) on satisfaction was mainly true for user's who talked to a chatbot. This finding would certainly be in line with the Expectation confirmation theory, which suggested that users' satisfaction increased specifically for the chatbot group because their expectations were lower. Thus, talking to a chatbot instead of a human really seems to amplify the positive effect of empathy on satisfaction (H3a), although not being statistically confirmed. The same was assumed for the user's trust (H3b) and loyalty (H3c). However, there was no evidence supporting these hypotheses. Only the results of the user's intention to use the service again, as one of the two subdimensions of loyalty, revealed rather striking outcomes. While empathy (vs. no empathy) greatly decreased the user's reuse intention when the interlocutor was a chatbot, no effect on the reuse intention could be identified in the human group. Many of the already discussed results indicate that users of healthcare services seem to be less sensitive to a human adviser showing or not showing empathy, while the reactions are more extreme when the adviser is a chatbot. A reason for that could lie in the few experiences that most people have made with chatbots. Nevertheless, it is only the fourth part of the analysis that yields the most diverse results and clearest answers, as it specifically investigates the affective and cognitive dimensions of empathy for the chatbot and human group.

This final analysis showed that affective empathy was substantially more appreciated from the human adviser, while cognitive empathy worked better when the adviser was a chatbot. This applied to all measured dependent variables. However, statistical significance could only be found for loyalty with the other variables only slightly missing the threshold. Hence, it can be confirmed that the user's interlocutor has a moderating effect. In particular, talking to a chatbot instead of a human weakens the positive effect of affective empathy compared to cognitive empathy on the user's loyalty to the service (H4c). Especially the user's word-of-mouth communications experiences this moderating effect. The same cannot be statistically confirmed but very well assumed for the user's satisfaction with (H4a) and trust in the service (H4b). Again, a possible explanation could be found in the individual expectations of the adviser, as well as the concept of genuineness. A chatbot trying to show affective empathy might be perceived as rather fake and less genuine as most people see computers as not being able to feel emotions (Nass & Moon, 2000). In contrast, cognitive empathy might be perceived as more believable and thus genuine. The opposite is the case when the adviser is human. Here, the user probably wishes for the other person to feel with them, while getting the perception of only being understood on a cognitive level could cause a very unpleasant experience. The latter is supported by the fact that the satisfaction, trust, and loyalty scores for users who were shown cognitive empathy by a human are the lowest scores of all.

In any case, these results provide several implications for theory, practice, as well as future research, which are presented in the following concluding chapter.

5. Conclusion

5.1 Theoretical and managerial implications

Theoretical implications. This study makes relevant contributions to the research fields of empathy in service and chatbots. In general, and without distinguishing between chatbot and human, it was confirmed once again that empathy plays a decisive role in the service context and greatly impact the user's perception of the experience. In this case the focus was on empathy from a healthcare provider's side, which was found to have an important influence on the user's satisfaction with, trust in, and loyalty to the service. In particular, satisfaction was mostly positively affected while trust and loyalty often seemed to even experience a minor decrease. Thus, this study's results seem to contradict previous literature suggesting an extremely close interrelation between these three constructs (Bahadur et al., 2019; Morgan & Hunt, 1994; Wieseke, Geigenmüller, & Kraus, 2012). However, a regression analysis could confirm a causal relationship between satisfaction and loyalty with trust acting as a mediator. Hence, this often-assumed interrelation appears to be valid in the context of chatbot healthcare services.

Still, this study's most meaningful theoretical contribution lies in the discovery that empathy works and takes effect differently in a chatbot compared to a human. In order for this difference to become visible, empathy must not be seen as a single but as the multidimensional construct that earlier research already suggested (Clark, Robertson, & Young, 2019; Cuff et al., 2016; Van der Graaff et al., 2016). Affective empathy and cognitive empathy should always be considered and investigated separately. This applies to general research about empathy, and particularly to research about the increasingly relevant technology of chatbots in healthcare, as they cannot be equated with humans. While a human generates more user satisfaction, trust and loyalty with affective empathy, a chatbot does the same with cognitive empathy. A possible explanation for this phenomenon could lie in the initially mentioned Expectation confirmation theory (Oliver, 1977). The different perceptions of what a programmed chatbot compared to a real human can and should do in a service setting, and especially what is perceived as either genuine or rather fake might majorly influence the user's pleasure of the experience.

Managerial implications. The few cases in this experiment, in which empathy had a negative effect on the user's trust or loyalty, were all missing statistical significance. Thus, much emphasis should still be placed on empathy in regular and healthcare services, which was again and again found to be extremely important (Wiseman, 1996; Reynolds, Scott, & Jessiman, 1999; Ioannidou & Konstantikaki, 2008). Nonetheless, developers of chatbots that are used by firms, hospitals and healthcare providers should know the fine differences of affective and cognitive empathy. Minor details can heavily affect a user's experience, but more importantly also a patient's well-being. Creating a chatbot that too frantically tries to simulate a real human by pretending to really feel the person's inner state might not be perceived as very natural and genuine, eventually leading to suffering people not reaching out anymore to the so crucial medical support. Hence, chatbots should be provided with a type of empathy that acknowledges their programmed nature, i.e., cognitive empathy. It seems like the best and most helpful chatbot is not the one that desperately tries to be a human, but the one that embraces being a chatbot.

5.2 Limitations and outlook

Despite making several valuable theoretical and practical contributions, this study also has its limitations and thus reveals opportunities for further research. First of all, it has to be noted that the effectiveness of the experiment's manipulation regarding the user's experienced type of empathy could not be confirmed. Although it can be assumed that this was mainly caused by the difficulty of precisely self-evaluating the perceived empathy by the respondents, future experiments should put a major focus on the hard task of accurately simulating affective and/or cognitive empathy.

Moreover, the fact that only 125 out of the total 197 had previous chatbot experience and could thus give answers for the control variables, resulted in the group sizes in the conducted ANCOVAs sometimes undercutting the needed threshold for statistical significance. This, in turn, led to many findings only being well supported tendencies but not statistically confirmed hypotheses. While most control variables did not show any main effects on satisfaction, trust, or loyalty, the user's chatbot acceptance constantly did. Particularly, chatbot acceptance showed a significant main effect on the user's trust in and loyalty to the service for those respondents who talked to a chatbot. It can thus be assumed that someone's general attitude towards and acceptance of this rather new technology heavily influences the user experience. Future studies about chatbots should consider including chatbot acceptance as a part of the conceptual model.

Another limitation can be identified in the study's missing investigation of simultaneous affective and cognitive empathy. As important as the separate examinations are, empathic human communication often consists of an affective and a cognitive part. Researching about the effects of co-existing affective and cognitive empathy in chatbots seems evident. The question will be whether the interplay of the two dimensions can lead to synergy effects and even make for a more pleasurable user experience than cognitive empathy alone. Furthermore, it still needs to be studied what role exactly the third dimension in the form of behavioral empathy plays in a chatbot setting.

Finally, the last remaining question is whether and how people's perception of chatbots and thus also of empathy in chatbots changes over time. An increasing number of application areas of chatbots and a potentially growing familiarity with and acceptance of the technology might result in evolving expectations and a changing sensitivity. The Expectation confirmation theory already suggests that the expected greatly impacts perception (Oliver, 1977). Therefore, clarifying answers can only be derived from further studies about this highly relevant construct of empathy in humanity's new interlocutor.

References

Adam, M., Wessel, M., & Benlian, A. (2020). AI-based chatbots in customer service and their effects on user compliance. *Electronic Markets*. https://doi.org/10.1007/s12525-020-00414-7

Albrecht, K. (2006). *Social Intelligence: The New Science of Success*. John Wiley & Sons.

Alsmadi, S. (2008). Marketing Research Ethics: Researcher's Obligations toward Human Subjects. *Journal of Academic Ethics*, *6*(2), 153-160. https://www.researchgate.net/deref/http%3A%2F%2Fdx.doi.org%2F10.1007%2Fs10805-008-9060-1

Aluttis, C., Bishaw, T., & Frank, M. W. (2014). The workforce for health in a globalized context - global shortages and international migration. *Global Health Action*, *7*(1). https://doi.org/10.3402/gha.v7.23611

Anders, G. (2017). *"Alexa, Understand Me"*. MIT Technology Review. https://www.technologyreview.com/2017/08/09/149815/alexa-understand-me/

Bahadur, W., Khan, A. N., Ali, A., & Usman, M. (2019). Investigating the Effect of Employee Empathy on Service Loyalty: The Mediating Role of Trust in and Satisfaction with a Service Employee. *Journal of Relationship Marketing, 19*(3), 229-252. https://doi.org/10.1080/15332667.2019.1688598

Bates, M. (2019). Health Care Chatbots Are Here to Help. *IEEE Pulse*, *10*(3), 12-14. https://ieeexplore-ieee-org.ru.idm.oclc.org/stamp/stamp.jsp?tp=&arnumber=8720289&tag=1

Batson, C. D., Chang, J., Orr, R., & Rowland, J. (2002). Empathy, Attitudes, and Action: Can Feeling for a Member of a Stigmatized Group Motivate One to Help the Group? *Personality and Social Psychology Bulletin, 28*(12), 1656–1666. https://doi.org/10.1177/014616702237647

Battineni, G., Chintalapudi, N., & Amenta, F. (2020). AI Chatbot Design during an Epidemic Like the Novel Coronavirus. *Healthcare, 8*(2), 154. https://doi-org.ru.idm.oclc.org/10.3390/healthcare8020154

Bernhardt, B. C., & Singer, T. (2012). The neural basis of empathy. *Annual Review of Neuroscience, 35*, 1–23. https://doi-org.ru.idm.oclc.org/10.1146/annurev-neuro-062111-150536

Bhirud, N., Tataale, S., Randive, S., & Nahar, S. (2019). A Literature Review on Chatbots in Healthcare Domain. *International Journal of Scientific & Technology Research, 8*(7), 225-231. https://www.researchgate.net/profile/Nivedita_Bhirud/publication/334836867_A_Literature_Review_On_Chatbots_In_Healthcare_Domain/links/5d42fd5392851cd04699a8f5/A-Literature-Review-On-Chatbots-In-Healthcare-Domain.pdf

Bickmore, T. W., Mitchell, S. E., Jack, B. W., Paasche-Orlow, M. K., Pfeifer, L. M., O'Donnell, J. (2010a). Response to a relational agent by hospital patients with depressive symptoms. *Interacting with Computers, 22*(4), 289–298, https://doi.org/10.1016/j.intcom.2009.12.001

Bickmore, T. W., Puskar, K., Schlenk, E. A., Pfeifer, L. M., & Sereika, S. M. (2010b). Maintaining reality: relational agents for antipsychotic medication adherence. *Interacting with Computers*, *22*(4), 276-288. https://doi-org.ru.idm.oclc.org/10.1016/j.intcom.2010.02.001

Bielen, F. (2007). Waiting time influence on the satisfaction-loyalty relationship in services. *Managing Service Quality, 17*(2), 174-193. https://doi.org/10.1108/09604520710735182

Bonoma, T. V. (1985). Case Research in Marketing: Opportunities, Problems, and a Process. *Journal of Marketing Research*, *22*(2), 199–208. https://doi.org/10.1177%2F002224378502200209

Bove, L. L. (2019). Empathy for service: benefits, unintended consequences, and future research agenda. *Journal of Services Marketing*, *33*(1), 31-43. https://doi-org.ru.idm.oclc.org/10.1108/JSM-10-2018-0289

Brandtzaeg, P. B., & Følstad A. (2017). Why People Use Chatbots. In: Kompatsiaris I. et al. (eds) Internet Science. INSCI 2017. Lecture Notes in Computer Science, vol 10673. Springer, Cham. https://doi.org/10.1007/978-3-319-70284-1_30

Busselle, R. W. (2001). Television Exposure, Perceived Realism, and Exemplar Accessibility in the Social Judgment Process. *Media Psychology*, *3*(1), 43–67. https://doi-org.ru.idm.oclc.org/10.1207/S1532785XMEP0301_03

Cameron, G., Cameron, D., Megaw, G., Bond, R., Mulvenna, M., O'Neill, S., Armour, C., & McTear, M. (2018a). Best Practices for Designing Chatbots in Mental Healthcare – A Case Study on iHelpr. In: Proceedings of the 32nd Human Computer Interaction Conference. http://dx.doi.org/10.14236/ewic/HCI2018.129

Cameron, G., Cameron, D., Megaw, G., Bond, R., Mulvenna, M., O'Neill, S., Armour, C., & McTear, M. (2018b). Back to the future: Lessons from Knowledge Engineering Methodologies for Chatbot Design and Development. In: Proceedings of the 32nd Human Computer Interaction Conference. http://dx.doi.org/10.14236/ewic/HCI2018.153

Carr, L., Iacoboni, M., Dubeau, M. C., Mazziotta, J. C., & Lenzi, G. L. (2003). Neural Mechanisms of Empathy in Humans: A Relay from Neural Systems for Imitation to Limbic Areas. *PNAS, 100*(9), 5497–5502. https://doi.org/10.1073/pnas.0935845100

Castonguay, L. G., Constantino, M. J., & Holtforth, M. G. (2006). The working alliance: Where are we and where should we go? *Psychotherapy: Theory, Research, Practice, Training*, *43*(3), 271–279. https://doi.org/10.1037/0033-3204.43.3.271

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology*, *76*(6), 893-910. https://psycnet.apa.org/doi/10.1037/0022-3514.76.6.893

Chartrand, T. L., & Lakin, J. L. (2013). The Antecedents and Consequences of Human Behavioral Mimicry. *Annual Review of Psychology*, *64*(1), 285–308. https://doi-org.ru.idm.oclc.org/10.1146/annurev-psych-113011-143754

Chaudhuri, A., & Holbrook, M. B. (2001). The Chain of Effects from Brand Trust and Brand Affect to Brand Performance: The Role of Brand Loyalty. *Journal of Marketing*, *65*(2), 81–93. https://doi-org.ru.idm.oclc.org/10.1509%2Fjmkg.65.2.81.18255

Chin, W. W., Johnson, N., & Schwarz, A. (2008). A fast form approach to measuring technology acceptance and other constructs. *Mis Quarterly*, *32*(4), 687–703. https://doi.org/10.2307/25148867

Christ, C. C., Carlo, G, & Stoltenberg, S. F. (2016). Oxytocin Receptor (OXTR) Single Nucleotide Polymorphisms Indirectly Predict Prosocial Behavior Through Perspective Taking and Empathic Concern. *Journal of Personality*, *84*(2), 204-213. https://doi.org/10.1111/jopy.12152

Clark, M. A., Robertson, M. M., & Young, S. (2019). "I feel your pain": A critical review of organizational research on empathy. *Journal of Organizational Behavior*, *40*(2), 166-192. https://doi.org/10.1002/job.2348

Cuff, B. M. P., Brown, S. J., Taylor, L., & Howat, D. J. (2016). Empathy: A review of the concept. Emotion Review, *8*(2), 144-153. https://doi.org/10.1177%2F1754073914558466

Dahiya, M. (2017). A tool of conversation: Chatbot. *International Journal of Computer Science and Engineering, 5*(5), 158-161. https://www.researchgate.net/publication/321864990_A_Tool_of_Conversation_Chatbot

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology, 44*(1), 113–126. https://doi.org/10.1037/0022-3514.44.1.113

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *Mis Quarterly, 13*(3), 319–340. https://doi.org/10.2307/249008

Davis, M. H. (2009). Empathy. In H. T. Reis, & S. Sprecher (Eds.), *Encyclopedia of human relationships* (pp. 516–520). Thousand Oaks, CA: Sage Publications, Inc. https://doi.org/10.4135/9781412958479.n167

Dawes, J. (2008). Do data characteristics change according to the number of scale points used? an experiment using 5-point, 7-point and 10-point scales. *International Journal of Market Research: The Journal of the Market Research Society, 50*(1), 61–78. https://doi-org.ru.idm.oclc.org/10.1177%2F147078530805000106

Decety, J. (2010). The Neurodevelopment of Empathy in Humans. *Developmental Neuroscience, 32*(4), 257-267. https://doi.org/10.1159/000317771

Decety, J., & Cowell, J. M. (2014). Friends or Foes: Is Empathy Necessary for Moral Behavior? *Perspectives on Psychological Science, 9*(5), 525–537. https://doi.org/10.1177/1745691614545130

Della Cava, M. (2016). *Microsoft CEO Nadella: "Bots are the new apps".* USA Today. https://eu.usatoday.com/story/tech/news/2016/03/30/microsof-ceo-nadella-bots-new-apps/82431672/

de Waal, F. B. (2008). Putting the altruism back into altruism: The evolution of empathy. *Annual Review of Psychology, 59*, 279–300. https://doi.org/10.1146/annurev.psych.59.103006.093625

Doney, P. M., & Cannon, J. P. (1997). An Examination of the Nature of Trust in Buyer-Seller Relationships. *Journal of Marketing, 61*(2), 35-51. https://doi.org/10.2307/1251829

Donio, J., Massari, P., & Passiante, G. (2006). Customer satisfaction and loyalty in a digital environment: An empirical test. *Journal of Consumer Marketing, 23*(7), 445-457. https://doi.org/10.1108/07363760610712993

Feng, J., Lazar, J., & Preece, J. (2004). Empathy and online interpersonal trust: A fragile relationship. *Behaviour & Information Technology*, *23*(2), 97-106. https://doi.org/10.1080/01449290310001659240

Field, A. P. (2018). *Discovering statistics using IBM SPSS statistics* (5th edition., Ser. Sage edge). SAGE publications.

Flückiger, C., Wampold, B. E., Del Re, A. C., & Symonds, D. (2012). How Central Is the Alliance in Psychotherapy? *A Multilevel Longitudinal Meta-Analysis. Journal of Counseling Psychology*, *59*(1), 10-17. https://doi.apa.org/doi/10.1037/a0025749

Franke, T., Attig, C., & Wessel, D. (2019). A Personal Resource for Technology Interaction: Development and Validation of the Affinity for Technology Interaction (ATI) Scale. *International Journal of Human-Computer Interaction*, *35*(6), 456–467. https://doi-org.ru.idm.oclc.org/10.1080/10447318.2018.1456150

Fry, E. B. (1963). *Teaching faster reading: a manual*. Cambridge: Cambridge University Press.

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind reading. *Trends in Cognitive Sciences*, *2*(12), 493–501. https://doi.org/10.1016/S1364-6613(98)01262-5

Gardiner, P. M., McCue, K. D., Negash, L. M., Cheng, T., White, L. F., Yinusa-Nyahkoon, L., Jack, B. W., & Bickmore, T. W. (2017). Engaging women with an embodied conversational agent to deliver mindfulness and lifestyle recommendations: a feasibility randomized control trial. *Patient Education and Counseling*, *100*(9), 1720–1729. https://doi-org.ru.idm.oclc.org/10.1016/j.pec.2017.04.015

Goadsby, P. J., Lipton, R. B., & Ferrari, M. D. (2002). Migraine - current understanding and treatment. *The New England Journal of Medicine*, *346*(4), 257–70. https://www.researchgate.net/deref/http%3A%2F%2Fdx.doi.org%2F10.1056%2FNEJMra010917

Gonzalez-Liencres, C., Shamay-Tsoory, S. G., & Brüne M. (2013). Towards a neuroscience of empathy: ontogeny, phylogeny, brain mechanisms, context and psychopathology. *Neuroscience and Biobehavioral Reviews*, *37*(8), 1537–1548. https://doi-org.ru.idm.oclc.org/10.1016/j.neubiorev.2013.05.001

Gonzalez, R. (2017). *Virtual Therapists Help Veterans Open Up About PTSD*. WIRED. https://www.wired.com/story/virtual-therapists-help-veterans-open-up-about-ptsd/

Gremler, D. D., & Brown, S. W. (1996). Service Loyalty: Its Nature, Importance, and Implications. In Edvardsson, B., Brown, S. W., Johnston, R. & Scheuing, E. E. (Eds.), *Proceedings American Marketing Association* (pp. 171-180).

Grudin, J., & Jacques, R. (2019). Chatbots, Humbots, and the Quest for Artificial General Intelligence. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Paper No. 209, 1-11. https://doi.org/10.1145/3290605.3300439

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th edition). Cengage.

Halpern, J. (2003). What is Clinical Empathy? *Journal of General Internal Medicine, 18*(8), 670–674. https://doi.org/10.1046/j.1525-1497.2003.21017.x

Hauser, D. J., Ellsworth, P. C., & Gonzalez, R. (2018). Are Manipulation Checks Necessary? *Frontiers in psychology, 9*: 998, 1-10. https://doi.org/10.3389/fpsyg.2018.00998

Hayes, A. F. (2018). *Introduction to Mediation, Moderation, and Conditional Process Analysis* (2nd edition). Guilford Press.

Hegel, F., Spexard, T., Vogt, T., Horstmann, G., & Wrede, B. (2006). Playing a different imitation game: Interaction with an Empathic Android Robot. In: Proc. 2006 IEEE-RAS International Conference on Humanoid Robots (Humanoids06), pp. 56–61. https://doi.org/10.1109/ICHR.2006.321363

Hernandez (2018). *Meet the Chatbots Providing Mental Health Care*. Wall Street Journal. https://www.wsj.com/articles/meet-the-chatbots-providing-mental-healthcare-1533828373

Hill, C. A. (1987). Affiliation Motivation: People Who Need People ... But in Different Ways. *Journal of Personality and Social Psychology, 52*(5), 1008–1018. https://doi.apa.org/doi/10.1037/0022-3514.52.5.1008

Hogan, R. (1969). Development of an Empathy Scale. *Journal of Consulting and Clinical Psychology, 33*(3), 307-316. https://psycnet.apa.org/doi/10.1037/h0027580

Homburg, C., Wieseke, J., & Bornemann, T. (2009). Implementing the Marketing Concept at the Employee-Customer Interface: The Role of Customer Need Knowledge. *Journal of Marketing, 73*(4), 64-81. https://doi.org/10.1509%2Fjmkg.73.4.064

Horvath, A. O., & Greenberg, L. S. (1989). Development and validation of the Working Alliance Inventory. *Journal of Counseling Psychology, 36*(2), 223-233. https://psycnet.apa.org/doi/10.1037/0022-0167.36.2.223

Hurlemann, R., Patin, A., Onur, O. A., Cohen, M. X., Baumgartner, T., Metzler, S., Dziobek, I., Gallinat, J., Wagner, M., Maier, W., & Kendrick, K. M. (2010). Oxytocin enhances amygdala- dependent, socially reinforced learning and emotional empathy in humans. *Journal of Neuroscience*, *30*(14), 4999–5007. https://doi.org/10.1523/JNEUROSCI.5538-09.2010

Ioannidou, F., & Konstantikaki, V. (2008). Empathy and Emotional Intelligence: What is it really about? *International Journal of Caring Sciences*, *1*(3), 118-123. http://www.internationaljournalofcaringsciences.org/docs/Vol1_Issue3_03_Ioannidou.pdf

Jackson, P., Meltzoff, A., & Decety, J. (2005). How do we perceive the pain of others? A window into the neural processes involved in empathy. *NeuroImage*, *24*(3), 771–779. https://doi.org/10.1016/j.neuroimage.2004.09.006

Juniper Research (2021). Digital voice assistants in use to triple to 8 billion by 2023, driven by smart home devices. https://www.juniperresearch.com/press/press-releases/digital-voice-assistants-in-use-to-8-million-2023

Kassim, N., & Abdullah, N. A. (2010). The effects of perceived service quality dimensions on customer satisfaction, trust, and loyalty in e-commerce settings. *Asia Pacific Journal of Marketing and Logistics*, *22*(3), 351-371. https://doi-org.ru.idm.oclc.org/10.1108/13555851011062269

Kim, S. S., Kaplowitz, S., & Johnston, M. V. (2004). The Effects of Physician Empathy on Patient Satisfaction and Compliance. *Evaluation & the Health Professions*, *27*(3), 237-251. https://doi-org.ru.idm.oclc.org/10.1177%2F0163278704267037

Leite, I., Pereira, A., Mascarenhas, S., Martinho, C., Prada, R., & Paiva, A. (2013). The influence of empathy in human-robot relations. *International Journal of Human-Computer Studies*, *71*(3), 250-260. https://doi.org/10.1016/j.ijhcs.2012.09.005

Leung, S.-O. (2011). A Comparison of Psychometric Properties and Normality in 4-, 5-, 6-, and 11-Point Likert Scales. *Journal of Social Service Research*, *37*(4), 412–421. https://doi-org.ru.idm.oclc.org/10.1080/01488376.2011.580697

Levenson, R. W., & Ruef, A. M. (1992). Empathy: A Physiological Substrate. *Journal of Personality and Social Psychology*, *63*(2), 234-246. https://psycnet-apa-org.ru.idm.oclc.org/doi/10.1037/0022-3514.63.2.234

Lord, S. P., Sheng, E., Imel, Z. E., Baer, J., & Atkins, D. C. (2015). More Than Reflections: Empathy in Motivational Interviewing Includes Language Style Synchrony Between Therapist and Client. *Behavior Therapy*, *46*(3), 296–303. https://doi.org/10.1016/j.beth.2014.11.002

Lucas, G. M., Gratch, J., King, A., & Morency, L. (2014). It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior*, *37*, 94-100. https://doi.org/10.1016/j.chb.2014.04.043

Ly, K. H., Ly, A., & Andersson, G. (2017). A fully automated conversational agent for promoting mental well-being: A pilot RCT using mixed methods. *Internet Interventions*, *10*, 39-46. https://doi.org/10.1016/j.invent.2017.10.002

Martin, D. J., Garske, J. P., & Davis, M. K. (2000). Relation of the therapeutic alliance with outcome and other variables: A meta-analytic review. *Journal of Consulting and Clinical Psychology, 68(*3), 438-450. https://doi.org/10.1037/0022-006X.68.3.438

Mayo Clinic (2020, September 22). *Migraine: Simple steps to head off the pain*. Mayo Clinic. Retrieved May 5, 2021, from https://www.mayoclinic.org/diseases-conditions/migraine-headache/in-depth/migraines/art-20047242

Miner, A. S., Milstein, A., Schueller, S., Hegde, R., Mangurian, C., & Linos, E. (2016). Smartphone-based conversational agents and responses to questions about mental health, interpersonal violence, and physical health. *Jama Internal Medicine, 176*(5), 619–25. https://doi-org.ru.idm.oclc.org/10.1001/jamainternmed.2016.0400

Mishra, J., & Morrissey, M. A. (1990). Trust in employee/employer relationships: A survey of West Michigan managers. *Public Personnel Management*, *19*(4), 443-486. https://psycnet.apa.org/doi/10.1177/009102609001900408

Moorman, C., Zaltman, G., & Deshpande, R. (1992). Relationships Between Providers and Users of Market Research: The Dynamics of Trust Within and Between Organizations. *Journal of Marketing Research*, *29*(3), 314-328. https://doi.org/10.1177%2F002224379202900303

Morgan, R. M., & Hunt, S. D. (1994). The Commitment-Trust Theory of Relationship Marketing. *Journal of Marketing*, *58*(3), 20-38. https://doi.org/10.2307/1252308

Nadarzynski, T., Miles, O., Cowie, A., & Ridge, D. (2019). Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: A mixed-methods study. *Digital Health*, *5*, 1-12. https://doi.org/10.1177%2F2055207619871808

Nakao, H., & Itakura, S. (2009). An Integrated View of Empathy: Psychology, Philosophy, and Neuroscience. Integrative Psychological and Behavioral Science, *43*(1), 42-52. https://doi.org/10.1007/s12124-008-9066-7

Nass, C., & Moon, Y. (2000). Machines and Mindlessness: Social Responses to Computers. Journal of Social Issues, *56*(1), 81-103. https://doi-org.ru.idm.oclc.org/10.1111/0022-4537.00153

Neururer, M., Schlögl, S., Brinkschulte, L., & Groth, A. (2018). Perceptions on Authenticity in Chat Bots. *Multimodal Technologies and Interaction*, *2*(3), 60. https://doi.org/10.3390/mti2030060

Nezlek, J. B., Feist, G. J., Wilson, F. C., & Plesko, R. M. (2001). Day-to-Day Variability in Empathy as a Function of Daily Events and Mood. *Journal of Research in Personality*, *35*(4), 401–423. https://doi.org/10.1006/jrpe.2001.2332

Oliver, R. L. (1977). Effect of Expectation and Disconfirmation on Postexposure Product Evaluations: An Alternative Interpretation. *Journal of Applied Psychology, 62*(4), 480-486. https://content.apa.org/doi/10.1037/0021-9010.62.4.480

Oliver, R. L. & Swan, J. E. (1989). "Consumer Perceptions of Interpersonal Equity and Satisfaction in Transactions: A Field Survey Approach," *Journal of Marketing*, *53*(2), 21-35. https://doi-org.ru.idm.oclc.org/10.2307/1251411

Orenstein, B. W. (2017, June 1). *7 Tips to Ease Migraine Pain*. Everyday Health. Retrieved May 5, 2021, from https://www.everydayhealth.com/hs/understanding-migraines/ease-migraine-pain/

Oxford Languages (2021). Definition: Satisfaction. In *Oxford English Dictionary*. Retrieved March 21, 2021, from https://global.oup.com/academic/product/the-oxford-english-dictionary-9780198611868?cc=gb&lang=en&

Pallant, Julie (2020). *SPSS survival manual. A step by step guide to data analysis using SPSS* (7th ed.). Routledge. https://doi.org/10.4324/9781003117452

Plank, R. E., Minton, A. P., & Reid, D. A. (1996). A short measure of perceived empathy. *Psychological Reports*, *79*(3), 1219-1226. https://doi.org/10.2466%2Fpr0.1996.79.3f.1219

Preston, S. D., & de Waal, F. B. M. (2002). Empathy: Its ultimate and proximate bases. *Behavioral and Brain Sciences*, *25*(1), 1–20. https://doi.org/10.1017/s0140525x02000018

Redmond. M. V. (1985). The relationship between perceived communication competence and perceived empathy. *Communication Monographs*, *52*(4), 377-382. https://doi.org/10.1080/03637758509376119

Redmond, M. V. (1989). The Function of Empathy (Decentering) in Human Relations. *Human Relations*, *42*(7), 593-605. https://doi.org/10.1177%2F001872678904200703

Reynolds, W. J., Scott, B., & Jessiman, W. C. (1999). Empathy has not been measured in client's terms or effectively taught: a review of the literature. *Journal of Advanced Nursing*, *30*(5), 1177-1185. https://doi.org/10.1046/j.1365-2648.1999.01191.x

Riek, L. D., Paul, P. C., & Robinson, P. (2010). When my robot smiles at me: Enabling human-robot rapport via real-time head gesture mimicry. *Journal on Multimodal User Interfaces, 3*(1-2), 99-108. https://doi-org.ru.idm.oclc.org/10.1007/s12193-009-0028-2

Roca, M., Torralva, T., Gleichgerrcht, E., Woolgar, A., Thompson, R., Duncan, J., & Manes, F. (2011). The role of Area 10 (BA10) in human multitasking and in social cognition: A lesion study. *Neuropsychologia*, *49*(13), 3525–3531. https://doi.org/10.1016/j.neuropsychologia.2011.09.003

Scherer, A., Wünderlich, N. V., & von Wangenheim, F. (2015). The Value of Self-Service: Long-Term Effects of Technology-Based Self-Service Usage on Customer Retention. *Mis Quarterly*, *39*(1), 177–200. https://www.jstor.org/stable/26628346

Schori, D. (2018, September 6). *Tipps gegen Migräne.* Helsana. Retrieved May 5, 2021, from https://www.helsana.ch/de/blog/themen/mit-krankheit-leben/tipps-gegen-migraene.html

Seltman, H. J. (2012). *Experimental Design and Analysis*. Carnegie Mellon University.

Shamay-Tsoory, S. G., Aharon-Peretz, J., & Perry, D. (2009). Two systems for empathy: a double dissociation between emotional and cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions. *Brain*, *132*(3), 617–627. https://doi-org.ru.idm.oclc.org/10.1093/brain/awn279

Shamay-Tsoory, S. G. (2011). The neural bases for empathy. *The Neuroscientist*, *17*(1), 18-24. https://doi.org/10.1177/1073858410379268

Stiff, J. B., Dillard, J. P., Somera, L., Kim, H., & Sleight, C. (1988). Empathy, communication, and prosocial behavior. *Communication Monographs*, *55*(2), 198–213. https://doi-org.ru.idm.oclc.org/10.1080/03637758809376166

Tielman, M. L., Neerincx, M. A., van, M. M., Franken, I., & Brinkman, W. P. (2017). How should a virtual agent present psychoeducation? influence of verbal and textual presentation on adherence. *Technology and Health Care: Official Journal of the European Society for Engineering and Medicine*, *25*(6), 1081–1096. https://doi-org.ru.idm.oclc.org/10.3233/THC-170899

Toomey, E. C., & Rudolph, C. W. (2017). Age-Conditional Effects in the Affective Arousal, Empathy, and Emotional Labor Linkage: Within-Person Evidence from an Experience Sampling Study. *Work, Aging and Retirement*, *4*(2), 145–160. https://doi-org.ru.idm.oclc.org/10.1093/workar/wax018

Tuan, L. T. (2012). Saigon eye hospital. *Leadership in Health Services*, *25*(2), 123–138. https://doi-org.ru.idm.oclc.org/10.1108/17511871211221046

Vachon, D. D., Lynam, D. R., & Johnson, J. A. (2014). The (non)relation between empathy and aggression: Surprising results from a meta-analysis. *Psychological Bulletin, 140*(3), 751-773. https://psycnet.apa.org/doi/10.1037/a0035236

Vaidyam, A. N., Wisniewski, H., Halamka, J. D., Kashavan, M. S., & Torous, J. B. (2019). Chatbots and conversational agents in mental health: a review of the psychiatric landscape. *Canadian Journal of Psychiatry. Revue Canadienne De Psychiatrie*, *64*(7), 456-464. https://doi-org.ru.idm.oclc.org/10.1177/0706743719828977

Van der Graaff, J., Meeus, W., de Wied, M., van Boxtel, A., van Lier, P. A. C., Koot, H. M., & Branje, S. (2016). Motor, affective and cognitive empathy in adolescence: Interrelations between facial electromyography and self-reported trait and state measures. *Cognition and Emotion*, *30*(4), 745–761. https://doi.org/10.1080/02699931.2015.1027665

van Dolen, W., de Ruyter, K., & Lemmink, J. (2004). An empirical assessment of the influence of customer emotions and contact employee performance on encounter and relationship satisfaction. *Journal of Business Research*, *57*(4), 437–444. https://doi-org.ru.idm.oclc.org/10.1016/S0148-2963(02)00277-1

Wahle, F., Kowatsch, T., Fleisch, E., Rufer, M., & Weidt, S. (2016). Mobile Sensing and Support for People With Depression: A Pilot Trial in the Wild. *JMIR mHealth and uHealth*, *4*(3), e111. https://doi-org.ru.idm.oclc.org/10.2196/mhealth.5960

Wahle, F., Bollhalder, L., Kowatsch, T., & Fleisch, E. (2017). Toward the Design of Evidence-Based Mental Health Information Systems for People With Depression: A Systematic Literature Review and Meta-Analysis. *Journal of Medical Internet Research*, *19*(5), e191. https://doi-org.ru.idm.oclc.org/10.2196/jmir.7381

Walter, H. (2012). Social Cognitive Neuroscience of Empathy: Concepts, Circuits, and Genes. *Emotion Review*, *4*(1), 9–17. https://doi.org/10.1177/1754073911421379

Wellman, H. M. (2014). *Making minds: How theory of mind develops*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199334919.001.0001

Wieseke, J., Geigenmüller A., & Kraus, F. (2012). On the Role of Empathy in Customer-Employee Interactions. *Journal of Service Research*, *15*(3), 316–331. https://doi-org.ru.idm.oclc.org/10.1177/1094670512439743

Wiseman, T. (1996). A concept analysis of empathy. *Journal of Advanced Nursing*, *23*(6), 1162-1167. https://doi.org/10.1046/j.1365-2648.1996.12213.x

Wispé, L. (1986). The distinction between sympathy and empathy: To call forth a concept, a word is needed. *Journal of Personality and Social Psychology*, *50*(2), 314-321.

Xu, A., Liu, Z., Guo, Y., Sinha, V., & Akkiraju, R. (2017). A New Chatbot for Customer Service on Social Media. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, pp. 3506-3510. https://doi.org/10.1145/3025453.3025496

Yin, R.K. (1994). *Case study research; design and methods*. Thousand Oaks (Ca): Sage.

Zaki, J., Weber, J., Bolger, N., & Ochsner, K. (2009). The neural bases of empathic accuracy. *Proceedings of the National Academy of Sciences*, *106*(27), 11382–11387. https://doi.org/10.1073/pnas.0902666106

Zeithaml, V. A., Berry, L. L., & Parasuraman, A. (1996). The Behavioral Consequences of Service Quality. *Journal of Marketing*, *60*(2), 31–46. https://doi-org.ru.idm.oclc.org/10.2307/1251929

Appendix

Appendix 1

Overview of measurement scales

Construct *(+ source of scale)*	**Definition**	**Scale items – original**	**Scale items – altered (chatbot interaction)**	**Scale items – altered (human interaction)**
Dependent variables				
Satisfaction with service *(Oliver & Swan, 1989)*	The fulfilment of one's wishes, expectations, or needs by the service, or the pleasure derived from the this	Please indicate how satisfied you were with your ____ by checking the space that best gives your answer. 1. displeased me / pleased me 2. disgusted me / contented me 3. very dissatisfied with / very satisfied with 4. did a poor job for me / did a good job for me 5. poor choice in buying from that ____ / wise choice in buying from that ____ 6. unhappy with / happy with 7. bad value / good value 8. frustrating / enjoyable 9. very unfavorable / very favorable **→ 7-point semantic scale**	Please indicate how satisfied you were with *the service* by checking the space that best gives your answer. 1. displeased me / pleased me 2. disgusted me / contented me 3. very dissatisfied with / very satisfied with 4. did a poor job for me / did a good job for me 5. poor choice in *using this service* / wise choice in *using this service* 6. unhappy with / happy with 7. bad value / good value 8. frustrating / enjoyable 9. very unfavorable / very favorable **→ 7-point semantic scale**	Please indicate how satisfied you were with *the service* by checking the space that best gives your answer. 1. displeased me / pleased me 2. disgusted me / contented me 3. very dissatisfied with / very satisfied with 4. did a poor job for me / did a good job for me 5. poor choice in *using this service* / wise choice in *using this service* 6. unhappy with / happy with 7. bad value / good value 8. frustrating / enjoyable 9. very unfavorable / very favorable **→ 7-point semantic scale**

<table>
<tr><td>Loyalty to service

(Zeithaml, Berry, & Parasuraman, 1996)</td><td>The degree to which a customer exhibits repeat usage behavior from a service provider, possesses a positive attitudinal disposition toward the provider, and considers using only this provider when a need for this service arises</td><td>1. Say positive things about XYZ to other people.
2. Recommend XYZ to someone who seeks your advice.
3. Encourage friends and relatives to do business with XYZ.

(→ Word-of-mouth communications)

4. Consider XYZ your first choice to buy ____ services.
5. Do more business with XYZ in the next few years.

(→ Intention to use service again)

→ 7-point likelihood scale</td><td>After using the service, please indicate how likely you are to do the following things.

1. I say positive things about this service to other people.
2. I recommend this service to someone who seeks my advice.
3. I encourage friends and relatives to use this service.

(→ Word-of-mouth communications)

4. I consider this service my first choice to get health-related advice.
5. I use this service more often in the future.

(→ Intention to use service again)

→ 7-point likelihood scale</td><td>After using the service, please indicate how likely you are to do the following things.

1. I say positive things about this service to other people.
2. I recommend this service to someone who seeks my advice.
3. I encourage friends and relatives to use this service.

(→ Word-of-mouth communications)

4. I consider this service my first choice to get health-related advice.
5. I use this service more often in the future.

(→ Intention to use service again)

→ 7-point likelihood scale</td></tr>
<tr><td>Trust in service

(Chaudhuri & Holbrook, 2001)</td><td>The willingness of the user to rely on the ability of the service to perform its stated function</td><td>1. I trust [brand].
2. I rely on [brand].
3. The [brand] brand is safe.
4. [Brand] is an honest brand.

→ 7-point Likert scale</td><td>Lastly, please indicate your agreement with the following statements regarding the service.

1. I trust this service.
2. I rely on this service.
3. This service is safe.
4. This service is honest.

→ 7-point Likert scale</td><td>Lastly, please indicate your agreement with the following statements regarding the service.

1. I trust this service.
2. I rely on this service.
3. This service is safe.
4. This service is honest.

→ 7-point Likert scale</td></tr>
</table>

Control variables				
Need for empathy *(Hill, 1987)*	An internal state of tension experienced as a discrepancy between the currently perceived empathy and the desired perceived empathy	1. If I feel unhappy or kind of depressed, I usually try to be around other people to make me feel better. 2. I usually have the greatest need to have other people around me when I feel upset about something. 3. One of my greatest sources of comfort when things get rough is being with other people. 4. When I have not done very well on something that is very important to me, I can get to feeling better simply by being around other people. 5. During times when I have to go through something painful, I usually find that having someone with me makes it less painful. 6. It seems like whenever something bad or disturbing happens to me, I often just want to be with a close, reliable friend. → **5-point Likert scale**	Next, please assess yourself by indicating your agreement with the following statements. *no alterations for items* → 7**-point Likert scale**	Next, please assess yourself by indicating your agreement with the following statements. *no alterations for items* → 7**-point Likert scale**
Satisfaction with previous chatbot interactions	The fulfilment of one's wishes, expectations, or needs by previous	Please indicate how satisfied you were with your ____ by checking	Now, please try to remember your previous interactions with chatbots. Then, please indicate how satisfied	Now, please try to remember your previous interactions with chatbots. Then, please indicate how satisfied

(Oliver & Swan, 1989)	chatbot interactions, or the pleasure derived from these	the space that best gives your answer. 1. displeased me / pleased me 2. disgusted me / contented me 3. very dissatisfied with / very satisfied with 4. did a poor job for me / did a good job for me 5. poor choice in buying from that ____ / wise choice in buying from that ____ 6. unhappy with / happy with 7. bad value / good value 8. frustrating / enjoyable 9. very unfavorable / very favorable **→ 7-point semantic scale**	you were with these interactions with the chatbots by checking the space that best gives your answer. 1. displeased me / pleased me 2. disgusted me / contented me 3. very dissatisfied with / very satisfied with 4. did a poor job for me / did a good job for me 5. poor choice in contacting this chatbot / wise choice in contacting this chatbot 6. unhappy with / happy with 7. bad value / good value 8. frustrating / enjoyable 9. very unfavorable / very favorable **→ 7-point semantic scale**	you were with these interactions with the chatbots by checking the space that best gives your answer. 1. displeased me / pleased me 2. disgusted me / contented me 3. very dissatisfied with / very satisfied with 4. did a poor job for me / did a good job for me 5. poor choice in contacting this chatbot / wise choice in contacting this chatbot 6. unhappy with / happy with 7. bad value / good value 8. frustrating / enjoyable 9. very unfavorable / very favorable **→ 7-point semantic scale**
Affinity for chatbot interaction *(Franke, Attig, & Wessel, 2019)*	Someone's tendency to actively engage in intensive chatbot interaction	1. I like to occupy myself in greater detail with technical systems. 2. I like testing the functions of new technical systems. 3. When I have a new technical system in front of me, I try it out intensively. 4. I enjoy spending time becoming acquainted with a new technical system.	Next, please indicate your agreement with the following statements by checking the space that best gives your answer. 1. I like to occupy myself in greater detail with *chatbots*. 2. I like testing the functions of *chatbots*. 3. When I have a *chatbot* in front of me, I try it out intensively. 4. I enjoy spending time becoming acquainted with a *chatbot.*	Next, please indicate your agreement with the following statements by checking the space that best gives your answer. 1. I like to occupy myself in greater detail with *chatbots*. 2. I like testing the functions of *chatbots*. 3. When I have a *chatbot* in front of me, I try it out intensively. 4. I enjoy spending time becoming acquainted with a *chatbot.*

		5. It is enough for me that a technical system works; I don’t care how or why. (reversed) 6. I try to make full use of the capabilities of a technical system. → **6-point Likert scale**	5. It is enough for me that a *chatbot* works; I don’t care how or why. (reversed) 6. I try to make full use of the capabilities of a *chatbot.* → **7-point Likert scale**	5. It is enough for me that a *chatbot* works; I don’t care how or why. (reversed) 6. I try to make full use of the capabilities of a *chatbot*. → **7-point Likert scale**
Chatbot acceptance *(Chin, Johnson, & Schwarz, 2008)*	Someone’s perception of ease of use and usefulness of a chatbot	To aid me in my (accomplishment of tasks), overall, I feel (system) as a (technology type) is: 1. inefficient / efficient 2. ineffective / effective 3. unhelpful / helpful 4. quite useless / quite useful 5. difficult to learn / easy to learn 6. difficult to manipulate / easy to manipulate 7. obscure to interact with / clear to interact with 8. difficult to master / easy to master → **9-point semantic scale**	Once again, please check the space that best gives your answer: Overall, I feel *a chatbot* is: 1. inefficient / efficient 2. ineffective / effective 3. unhelpful / helpful 4. quite useless / quite useful 5. difficult to learn / easy to learn 6. difficult to manipulate / easy to manipulate 7. obscure to interact with / clear to interact with 8. difficult to master / easy to master → **7-point semantic scale**	Once again, please check the space that best gives your answer: Overall, I feel *a chatbot* is: 1. inefficient / efficient 2. ineffective / effective 3. unhelpful / helpful 4. quite useless / quite useful 5. difficult to learn / easy to learn 6. difficult to manipulate / easy to manipulate 7. obscure to interact with / clear to interact with 8. difficult to master / easy to master → **7-point semantic scale**
Manipulation checks				
Perception of interlocutor *(self-developed)*	The user’s perception of their interlocuter being a chatbot or a human	Now please think of your dialog partner in the conversation at the start of this survey. While imagining this conversation was	*no alterations for items*	*no alterations for items*

		real, please indicate your agreement with the following statements. 1. My dialog partner is a human being. 2. My dialog partner is a programmed chatbot. (reversed) 3. My dialog partner is a real person. **→ 7-point Likert scale**	**→ 7-point Likert scale**	**→ 7-point Likert scale**
Perceived affective empathy *(Plank, Minton, & Reid, 1996)*	The perception of someone else experiencing an affective state that is congruent with the own affective state	1. I have lousy feelings when dealing with this salesperson. (reversed) 2. I feel as if I am on the same wavelength as this salesperson. 3. This salesperson has a lot of knowledge about how I need to make decisions. 4. This salesperson seemed to feel what I needed when we talked about my purchase. **→ 5-point Likert scale**	Once again, please think of the chatbot with whom you had the conversation at the start of this survey. Then indicate your agreement with the following statements. 1. I have lousy feelings when dealing with this *chatbot*. (reversed) 2. I feel as if I am on the same wavelength as this *chatbot*. 3. This *chatbot* has a lot of knowledge about how I need to make decisions. 4. This *chatbot* seemed to feel what I needed when we talked about my *health concern*. **→ 7-point Likert scale**	Once again, please think of the healthcare professional with whom you had the conversation at the start of this survey. Then indicate your agreement with the following statements. 1. I have lousy feelings when dealing with this person. (reversed) 2. I feel as if I am on the same wavelength as this *person*. 3. This *person* has a lot of knowledge about how I need to make decisions. 4. This *person* seemed to feel what I needed when we talked about my *health concern*. **→ 7-point Likert scale**

Perceived cognitive empathy *(Plank, Minton, & Reid, 1996)*	The perception of someone else understanding the own internal state (i.e., thoughts and affective state)	1. This salesperson understands me and my role in this organization. 2. This salesperson really understood my feelings about this situation. 3. This salesperson does not understand how I think. (reversed) 4. This salesperson always understood our company's needs. → **5-point Likert scale**	1. This *chatbot* understands me and my *situation*. 2. This *chatbot* really understood my feelings about this situation. 3. This *chatbot* does not understand how I think. (reversed) 4. This *chatbot* always understood *my* needs. → **7-point Likert scale**	1. This *person* understands me and my *situation*. 2. This *person* really understood my feelings about this situation. 3. This *person* does not understand how I think. (reversed) 4. This *person* always understood *my* needs. → **7-point Likert scale**
Perceived realism *(Busselle, 2001)*	The user's judgement of the degree to which the narrative world is reflective of the real world	1. The crime you see on TV crime shows is very similar to crime in real life. 2. If I were to go to a hospital, I would not expect it to be like the hospitals I see on television. (reversed) 3. Characters in drama programs, like Beverly Hills 90210 or Melrose Place, are very similar to people in the real world. 4. The romantic relationships portrayed in drama programs are not at all like romantic relationships in the real world. (reversed)	Lastly, please think of the whole situation that was described to you at the beginning of this survey, including the conversation you had with the chatbot. Then please indicate your agreement with the following statements. 1. The *situation* is very similar to *situations* in real life. 2. If I were to *have a health concern and seek advice*, I would not expect it to be like *it is described here*. (reversed) 3. *The situation is* very similar to *situations* in the real world.	Lastly, please think of the whole situation that was described to you at the beginning of this survey, including the conversation you had with the healthcare professional. Then please indicate your agreement with the following statements. 1. The *situation* is very similar to *situations* in real life. 2. If I were to *have a health concern and seek advice*, I would not expect it to be like *it is described here*. (reversed) 3. *The situation is* very similar to *situations* in the real world.

		→ **7-point scale**	4. The *situation* portrayed *here is* not at all like *a situation* in the real world. (reversed) → **7-point Likert-scale**	4. The *situation* portrayed *here is* not at all like *a situation* in the real world. (reversed) → **7-point Likert-scale**