Radboud University

Nijmegen, the Netherlands

Unraveling chatbot confusion

Exploring the importance of the sensorial and behavioral dimension of the customer experience in the relationship between customer confusion and customer satisfaction and loyalty

Master thesis

MSc in Business Administration: Marketing specialization

Radboud University

30 June 2023

Author: Annick Evers

Student number: S1082464

Thesis supervisor: Ruud Wetzels

Second examiner: Karim Sidaoui

Abstract

Purpose – To investigate how customers experience confusion in the context of chatbots, by examining the effect of confusion (and the interaction with time pressure) on the sensorial and behavioral dimension of the customer experience (CE). Additionally, the effects of confusion on CE-outcomes such as satisfaction and loyalty are examined, also via the CE-dimensions.
Design/methodology – This research uses a 2 x 2 between-subjects experimental design with confusion and time pressure being manipulated. During a shopping experience with an eye-tracker and a shopping bot, 200 participants answered closed and open-ended questions, making this research a mixed-methods study. Structural Equation Modelling was used to analyze the measurement and structural model and examine the hypothesized effects.
Findings – Findings demonstrate that the relationship between confusion and CE-outcomes is only mediated by the behavioral dimension; the sensorial dimension had no mediation effect. Additionally, a negative effect from confusion on satisfaction was found, as well as a positive effect from confusion on the sensorial and behavioral dimension.
Research implications – This research contributes to existing literature by empirically measuring the CE-dimensions (and their interrelationships) by using attitudinal and behavioral data, therefore broadening and deepening the understanding of this concept. Furthermore, confusion is studied from a more holistic view by examining the effect on sensations and behavior.
Managerial implications – As reducing and preventing confusion is crucial for firms, this research illustrates that managers are able to recognize confusion in an early stage and can adjust the webpage accordingly. Additionally, this research shows that preventing confusion starts with providing clear, unambiguous information, in the right quantity, to the corresponding target group.
Keywords – Chatbots, Eye-tracking, Confusion, Customer experience, Sensorial dimension, Behavioral dimension, Satisfaction, Loyalty

Table of content

1. Introduction 5

2. Theoretical background 10

2.1 Self-service technologies 10

2.1.1 Chatbots 10

2.2 Customer experience 12

2.2.1 The interrelations between the customer experience dimensions 14

2.3 Sensorial processing 15

2.3.1 Vision and visual processing 15

2.4 The behavioral dimension 18

2.5 Hypotheses development 19

2.5.1 Interrelation between the sensorial dimension and behavioral dimension 19

2.5.2 Drivers of customer experience 22

2.5.3 Outcomes of the customer experience 25

2.5.4 Serial mediation and mediation 27

3. Methodology 29

3.1 Design 29

3.2 Measurement 30

3.2.1 Measurement of dynamic gaze behavior 30

3.2.2 Measurement of other constructs 31

3.3 Procedure 33

3.4 Data analysis procedure 33

3.5 Participants 34

3.6 Research ethics 34

4. Research results 35

4.1 Data preparation 35

4.1.1 Cleaning the data and manipulation check 35

4.1.2 Missing data analysis 37

4.2 Evaluation of the measurement model *37*

4.3 Evaluation of the structural model *39*

4.3.1 Collinearity and Coefficient of Determination 39

4.3.2 Effect sizes 39

4.3.3 Path coefficients 40

4.3.4 Additional analysis 42

Control variables *43*

5. Conclusion and discussion 44

5.1 Discussion *44*

5.2 Theoretical implications *49*

5.3 Practical and managerial implications *51*

5.4 Limitations and further research *54*

References 55

Appendices 70

Appendix 1. Visualization of the human eye *70*

Appendix 2. Design of the experiment *71*

Appendix 3. Manipulation check questions *74*

Appendix 4. Operationalization table *75*

Appendix 5. Full script *77*

Appendix 6. Consent form *83*

Appendix 7. Results manipulation check pre-test *84*

Appendix 7A: Confirmatory factor analysis 84

Appendix 7B: Manipulation check pre-test with summated scales 89

Appendix 8. Results manipulation check experiment *92*

Appendix 8A: Confirmatory factor analysis 92

Appendix 8B: Manipulation check experiment with summated scales 96

Appendix 9. Results *99*

Appendix 9A: Data preparation 99

Appendix 9B: Evaluation of the measurement model 101
Appendix 9C: Evaluation of the structural model 111
Appendix 9D: Additional qualitative analysis behavior 121

1. Introduction

The era of rapid technological growth has radically transformed the opportunities for firms to interact with their customers and provide customers with information. Firms are extensively using self-service technologies (SSTs) to interact with customers and improve customer service (Chan & Petrikat, 2022). Self-service technologies are technologies that can be used independently without interaction or assistance from an employee (Marfo et al., 2022). These SSTs are changing the way a customer is experiencing the interaction with the firm (Marfo et al., 2022). SSTs provide benefits for the firm, including competitive advantage, efficiency, effectiveness, and enhanced financial performance (Chan & Petrikat, 2022). One of those self-service technologies are chatbots (Sheehan et al., 2020). A chatbot is an artificially intelligent program that is able to interact with customers (Riikkinen et al., 2018) In recent years, chatbots have emerged as a means to provide customer service activities, intending to increase customer satisfaction (Nicolescu & Tudorache, 2022). The usage of chatbots for firms is generally motivated by reasons such as cost reduction, time-saving, and serving customers simultaneously. For customers, chatbots offer benefits such as having 24/7 access to customer service which enables customers to ask any question at any time, (Nicolescu & Tudorache, 2022; Adamopoulou & Moussiades, 2020; Fotheringham & Wiles, 2022), improving the experience through personalization (Fotheringham & Wiles, 2022), and increasing convenience (Nicolescu & Tudorache, 2022). In 2021, the chatbot market was valued at 3.49 billion U.S. dollars and is forecasted to reach 22.9 billion U.S. dollars by 2030 (Businesswire, 2022). In addition, 1.4 billion people use chatbots nowadays and this number keeps on growing (Beckmann, 2022). On average, chatbots can increase sales by 67% (Beckmann, 2022).

Practical need

Even though digital and artificial intelligent-driven options offer benefits for customers, there is also a dark side to this. A risk with chatbots is that the customer may not fully understand the message they receive. This results in a greater chance for customers to get confused while consulting chatbots (Lu & Gursoy, 2014), and thus the interaction with a chatbot is experienced as negative. This unfavorable experience negatively affects the satisfaction of the customer (Ma et al., 2023; Nicolescu & Tudorache, 2022). Confusion caused by chatbots has major disadvantages for firms. The 67% increase in sales that a chatbot can cause, will drop drastically when confusion occurs. First, as Businesswire (2023) states, almost half of the

people interacting with a chatbot receive answers from chatbots that are invaluable. As a result, the negative experience with the chatbot leads to customers purchasing another brand, abandoning the purchase, or telling friends and family about it in 30% of the cases. This is problematic for firms because 92% of the customers believe recommendations from friends and family before believing other forms of advertising (Whitler, 2014). In line with this, Edward and Sahadev (2012) argue that confusion and the negative experience resulting from this can cause dissatisfaction with a customer. Second, confusion can cause a decrease in customer loyalty which means that customers have lower repeat visit intentions and lower (repeat) purchases. Third, confusion leads to a poorer brand image and more returned products (Mitchell & Papavassiliou, 1999; Edward & Sahadev, 2012; Garaus & Wagner, 2016). Additionally, Fitzgerald et al. (2019) argue that customer confusion reduces the perception of product quality and damages firm credibility. These disadvantages undermine the potential 67% increase in sales that chatbots can bring. All these disadvantages of confusion eventually lead to decreased profits (Walsh et al., 2007). On top of that, firms spend over 1.3 trillion U.S. dollars per year to address customer requests (Zabój, 2022). When a chatbot is therefore not functioning as it was intended, both the sales are lost and the investments in customer service are worthless. It is therefore crucial for firms to understand confusion in the context of chatbots and how customers experience this to enable them to resolve it and to prevent a firm from losing sales, customers, credibility, investments, and profit. However, existing research on confusion unfortunately offers little guidance.

Theoretical need

Existing research on customer confusion is more focused on its antecedents and consequences. For example, Matzler et al. (2011) find that product knowledge and usability have a significant, negative impact on confusion. Several (more recent) studies have focused on the consequences of confusion. Shiu (2017), Lu and Gursoy (2014), and Sharma et al. (2022) found that confusion has an impact on postponing or abandoning a decision or purchase. Moreover, the impact of confusion on outcomes such as trust, satisfaction, word of mouth, and loyalty has been extensively studied, both in the past as well as in present research (Moon et al., 2016; Matzler et al., 2011; Tjiptono et al., 2014; Edward & Sahadev, 2012; Walsh & Mitchell, 2010; Mitchell et al., 2005; Walsh et al., 2007). However, measures of confusion are oversimplified in academic literature (Shiu, 2017; Edward & Sahadev, 2012; Walsh & Mitchell, 2010). Therefore, a more holistic understanding of how people experience confusion is needed.

To meet this necessity, the customer experience offers a theoretical lens. The customer experience is increasingly becoming a dominant topic for academics. It is defined as "encompassing every aspect of a company's offering—the quality of customer care, of course, but also advertising, packaging, product and service features, ease of use, and reliability. It is the internal and subjective response customers have to any direct or indirect contact with a company" (Meyer and Schwager, 2007, p. 1). The customer experience is an important concept to evaluate a firm's performance. Providing a positive customer experience is essential for achieving satisfaction, competitive advantage, and loyalty (Jain et al., 2017; Hwang & Seo, 2016). However, measuring the customer experience is complex because of its personal and subjective nature, the lack of a clear definition, and the multidimensionality (Jain et al., 2017). The customer experience encompasses five separate dimensions: sensorial, emotional, cognitive, behavioral, and social (Schmitt, 1999; Verhoef et al., 2009; Brakus et al., 2009). The sensorial dimension encompasses the senses; the emotional relates to affect, mood, and feelings; the cognitive dimension includes the mental processes; the behavioral dimension involves (physical) actions; the social dimension encompasses social relationships (Gentile et al., 2007; Keiningham et al., 2017; Brun et al., 2017).

In literature, academics call for more research and data on (a) how the customer experience can be empirically measured by behavioral data, taking into account its multidimensional nature (Jain et al., 2017; Hwang & Seo, 2016; Lemon & Verhoef, 2016), (b) the drivers and consequences of the customer experience (Lemon & Verhoef, 2016), and (c) the sensorial and behavioral dimension of the customer experience and how these interrelate (Mahr et al., 2019; Gerea et al., 2021; Lemon & Verhoef, 2016).

Research objective

In this research, the following research question is answered: "How is the relationship between customer confusion and customer satisfaction and loyalty mediated by the sensorial and behavioral dimension of the customer experience and how is this moderated by perceived time pressure?". By investigating this, the afore-mentioned gaps are addressed by focusing on (a) the multidimensionality of the customer experience and how it can be empirically measured using behavioral data (b) confusion as a driver of the customer experience and customer satisfaction and loyalty as consequences of the customer experience, and (c) the sensorial and behavioral dimension and how these relate to each other. Especially the sensorial dimension requires more attention (Mahr et al., 2019).

The sensorial dimension encompasses the five senses: sight, touch, sound, taste, and smell. Since this research is conducted in the context of chatbots, the visual aspect of the senses is most important. An individual visually processes an environment through two factors; the bottom-up factors and the top-down factors. Bottom-up factors are stimulus-driven processes (Orquin et al., 2013) and "refer to stimulus parts that extract fixations independent of the internal state of an observer" (Schütt et al., 2019, p. 2). This entails that bottom-up factors start in the external environment. Anninou and Foxall (2019) argue that confusion is a response to the environment and the conditions present in that environment. Therefore, it can be argued that confusion is a bottom-up factor. In addition, top-down factors are goal-driven processes (Orquin et al., 2013) and "refer to cognitive influences on the chosen fixation locations based on the current aims of an observer varying with task demands and memory" (Schütt et al., 2019, p. 2). This means that top-down factors start in the brain with a certain goal or aim in mind. When performing a task with a specific goal in mind, perceived time pressure plays a role. Perceived time pressure is a subjective feeling or thought that is determined by the brain (Maruping et al., 2015; Ordóñez et al., 2015; Krishna, 2013; Agapito et al., 2013). These two factors entail that gaze behavior is determined by the interplay of bottom-up (confusion) and top-down factors (perceived time pressure) (Orquin & Mueller Loose, 2013).

Since the sensorial dimension requires more attention, the purpose of this research is to examine the sensorial and behavioral dimensions of the customer experience to understand how customers experience confusion by focusing on dynamic gaze behavior and searching for additional information, and response time, thereby addressing the first gap. Furthermore, this research examines how these dimensions of the customer experience mediate the relationship between customer confusion and customer satisfaction and loyalty, thereby addressing the second gap and taking a more holistic approach on studying confusion. Last, this research empirically investigates how people visually process information and how this impacts their behavior. In other words, the interrelation between the sensorial dimension and the behavioral dimension is empirically examined, thereby addressing the third gap and incorporating data on the senses and behaviors of individuals.

Relevance

Drawing on eye-tracking data from 200 participants to investigate the dynamic gaze behavior, and a lab experiment with behavioral data, this research makes important contributions in the context of chatbots. First, this research contributes to filling the gap of empirically measuring

the customer experience and its dimensions by using attitudinal and behavioral data the examine the dimensions and the outcomes of the customer experience. By using an eye-tracker to study sensations and by using real-time behavioral data to study behavior, this research shows how the dimensions can be empirically measured in a chatbot context. Second, this research contributes to the literature by taking a more holistic view of studying confusion by examining how customers experience it through their senses and their behavior. In contrast to other studies (e.g., Shiu, 2017), this research takes a more holistic perspective by studying the effect of confusion on dynamic gaze behavior, on searching for additional information, and on response time to examine how individuals experience confusion. Third, the sensorial and behavioral dimensions of the customer experience are not studied in isolation, but in terms of how they are related, which is not fully understood yet in literature. By means of a combination of eye-tracking and chatbots, this research shows that the dimensions of the customer experience can be measured simultaneously, taking into account their interrelations. Due to this unique combination of data regarding the sensorial and behavioral dimension, and the context of chatbots, this research is one of the first to measure confusion in this way.

The results provide insights for managers into the gaze behavior and actual behavior of an individual when encountering confusion. By means of the results, managers are able to recognize confusion in an early stage of the experience, thereby having the ability to respond to this confusion by providing additional information or by adjusting the webpages accordingly. This is crucial for firms since confusion leads to major disadvantages when this is not resolved. Additionally, the results demonstrate that managers can prevent confusion by providing correct, clear, unambiguous information, and the right amount of information to the corresponding target group, by means of a chatbot.

Outline

The remainder of this research continues as follows. First, the theoretical background of the conceptual model is discussed, as well as the corresponding hypotheses. Then, the methodology of the research is clarified. Next, the results are addressed, followed by the conclusion and discussion. Last, theoretical and practical implications, limitations, and recommendations for further research are addressed.

2. Theoretical background

2.1 Self-service technologies

Due to advances in technology, the traditional way of service delivery is substituted by developments in technology (Lee & Yang, 2013). A widely adopted technology to deliver service and interact with customers is self-service technologies (SSTs) (Chan & Petrikat, 2022). Self-service technologies are defined as "technologies that customers use independently without interaction or assistance from a firm's employees" (Marfo et al, 2022, p. 1). Examples of SSTs include ATMs (Chen et al., 2021), self-check-in at airports, and automated chatbots (Chan & Petrikat, 2022). SSTs offer benefits for firms such as reducing labor costs, creating competitive advantage, improving efficiency and effectiveness, and enhancing financial performance (Chan & Petrikat, 2022; Oh et al., 2013). SSTs can be considered as a replacement for employees when the SST is fully automated (Chen et al., 2021). Due to this replacement of personal and face-to-face interaction, the way a customer experiences the interaction and the service has changed (Marfo et al., 2022). This will be further elaborated on in section 2.3. When the experience with an SST is favorable, SSTs offer several benefits to customers such as the opportunity to choose how and when they would like to receive a service, cheaper transactions, the ability to co-create value, and the service is more customized and personalized (Kokkinou & Cranage, 2013).

Despite the many advantages SSTs offer, there are also challenges involved. First, resistance to SSTs is a common problem (Chan & Petrikat, 2022). A reason for this might be that the SST is too complicated or that past experiences were unfavorable (Chan & Petrikat, 2022). Second, the technology needs to be well-developed, adaptable to changes, accurate and useful, safe to use regarding personal information, and easy to use. A third and last challenge is that implementing SSTs is expensive and uncertain. Investments need to be made while it is uncertain whether these investments will be worth it or not (Chan & Petrikat, 2022).

2.1.1 Chatbots

One of the examples of SSTs mentioned in the section above are chatbots. A chatbot is a combination of a 'chat', and a 'bot'. They rely on artificial intelligence and computational linguistics techniques and due to this, the chatbot is able to understand natural language and respond in natural language to customers' requests (Rese et al., 2020). Chatbots are defined as "an artificially intelligent program that can interact with consumers via different messaging apps" (Riikkinen et al., 2018, p. 1148). Chatbots are primarily designed to increase customer

satisfaction by providing customer service activities, sales/marketing activities, and processing orders. The chatbot performing these activities can be used along the entire customer journey (prepurchase, purchase, and post-purchase). For example, a chatbot can assist in the prepurchase stage by applying predictive modeling and algorithms to recommend a product based on needs, expectations, and desired price points (Rese et al., 2020).

The use of chatbots by firms is motivated by reasons such as cost reduction, time-saving, and the ability to serve customers simultaneously (Nicolescu & Tudorache, 2022). Chatbots also are beneficial to customers in terms of having 24/7 access to customer service (Nicolescu & Tudorache, 2022; Adamopoulou & Moussiades, 2020; Fotheringham & Wiles, 2022), improving the experience through personalization (Fotheringham & Wiles, 2022), and increasing convenience (Nicolescu & Tudorache, 2022). However, despite these benefits, customers are sometimes reluctant to use chatbots due to the immaturity of the technology, the lack of personal interaction with a service employee, the chatbot being inaccurate or unclear, and the feeling of being observed when using a chatbot (Rese et al., 2022). When the chatbot is not working as intended or is responding in an unclear way, a risk that emerges is that a customer gets confused (Lu & Gursoy, 2014). A confused customer will negatively experience the interaction with a chatbot. As a result, confusion will negatively affect the satisfaction and loyalty of the customer (Ma et al., 2023; Nicolescu & Tudorache, 2022). To gain more insight into confusion, the subsequent section elaborates on this concept.

2.2 Confusion

Confusion is a broad concept defined differently by several studies. Turnbull et al. (2000, p. 145) define confusion as “consumer failure to develop a correct interpretation of various facets of a product/service, during the information processing procedure”. Edward and Sahadev (2012, p. 128) define customer confusion as “an uncomfortable psychological state consumers experience when exposed to an overload of marketing information which is often very similar, misleading, ambiguous and inadequate in nature”. Shukla et al. (2010, p. 292) define confusion as “a mental state characterized by a lack of clear and orderly thought and behavior”. Moreover, based on the research of Mitchell et al. (2005) confusion leads to more difficulty with selecting, interpreting, and evaluating stimuli. Since there is no consensus in literature, this research proposes the following description: “Confusion is an uncomfortable psychological state customers experience when exposed to an overload of marketing information which is often very similar, misleading, ambiguous and inadequate in nature. This psychological state affects behavioral responses and thoughts since the customer is

unable to correctly select, interpret, and evaluate various facets of a product or service during the information processing procedure" (Turnbull et al, 2000; Mitchell & Papavassiliou, 1999; Edward & Sahadev, 2012; Leek & Kun, 2006; Mitchell et al., 2005).

As Meuter et al. (2005) and Lu and Gursoy (2014) discuss, when an SST is complex, unclear, or inaccurate, the interaction with it and its usage will be challenging, which might cause confusion. As a result of the confusion that follows, the benefits mentioned in section 2.1 are becoming less apparent to the customer. Additionally, this confusion will influence the adoption process of new SSTs, and customers will be more reluctant to try these new technologies due to the negative experience (Meuter et al., 2005). Looking at the proposed description of confusion mentioned above, it can be concluded that confusion is a state that is experienced by individuals and affects the thoughts and behavior of individuals. Therefore, the customer experience, which consists of several dimensions including thoughts and behavior, is an essential concept to elaborate on.

2.2 Customer experience

Several authors (Schmitt, 1999; Verhoef et al., 2009; Brakus et al., 2009) state that the customer experience encompasses five separate experiences: sensorial, emotional, cognitive, behavioral, and social. The customer experience is therefore multidimensional and holistic in nature (Lemon & Verhoef, 2016). First, the sensorial dimension consists of the experience related to the five senses: sight, touch, sound, taste, and smell. These senses are activated by a certain stimulus, mainly to arouse excitement, satisfaction, aesthetical pleasure, and/or a sense of beauty (Gentile et al., 2007). Second, the emotional dimension is related to the affective system which generates moods, feelings, and emotions. The emotional experience is mainly focused on creating an emotional bond with a firm, brand, or product (Gentile et al., 2007). Third, the cognitive dimension is connected with conscious mental (thinking) processes. The cognitive processes are related to memory, perception, problem-solving, language, and abstract thinking (Keiningham et al., 2017). Fourth, the behavioral dimension affects the customer physically. This entails that the experience can modify an action taken, or it can change certain habits or a lifestyle. The action taken is related to the consumption of a product or service (Brun et al., 2017). Fifth and last, the social dimension relates to the individual his/herself and the relationships that he/she has with other people. This dimension is addressed by encouraging individuals to use a product or service with other people or to affirm his/her social identity (Gentile et al., 2007).

Due to this multidimensionality of the customer experience, the interaction between firm and customer is more important than ever. Specifically, the interaction between firms and customers is characterized by an excessive number of touchpoints, which makes a favorable customer experience essential. However, creating, managing, and attempting to control the customer experience has become very complex nowadays (Lemon & Verhoef, 2016). However, the customer experience is important to evaluate a firm's performance because the customer experience is important for achieving differentiation and sustainable competitive advantage. Additionally, customer satisfaction, loyalty, and word of mouth are positively impacted by the customer experience (Jain et al., 2017). As a result, the customer experience has been studied extensively and is broadly defined by Meyer and Schwager (2007, p. 1) as "encompassing every aspect of a company's offering—the quality of customer care, of course, but also advertising, packaging, product and service features, ease of use, and reliability. It is the internal and subjective response customers have to any direct or indirect contact with a company". An experience is perceived differently by every customer, which is in line with the shift from commoditization to personalization (Jain et al., 2016).

The customer experience has changed due to the increasing number of self-service technologies and the replacement of face-to-face interaction resulting from this. According to Åkesson et al. (2014), a customer co-creates the experience it has with the service system. The service system, the drivers of the customer experience, and the customer experience itself are interdependent when using an SST. Åkesson et al. (2014) distinguish four different types of drivers that influence the customer experience when interacting with an SST. The first type is informational drivers (being inspired, obtaining information, evaluating alternatives). The second type is relational drivers (reliable use, convenient, self-controlling). The third type of organizational drivers (service orientation, problem-solving activity, accessibility and availability, (lack of) trust). The fourth and last type is technological drivers (perceived capacity, ease of use, flexibility) (Åkesson et al., 2014). The above-mentioned drivers in the four categories determine whether an interaction with an SST is experienced as favorable or unfavorable. For example, when customers are interacting with an SST and they feel as if they are controlled or overlooked (lack of trust), this will negatively affect the experience with the SST (Åkesson et al., 2014). All the drivers can lead to a favorable experience while interacting with an SST when the SST is properly designed.

When interacting with an SST, customers feel a certain way about this interaction, they think about the interaction, and this eventually impacts the (future) behavior (Meuter et al., 2000). Due to this multidimensionality and the subjectivity of the customer experience, it

is a complex concept to define, conceptualize, and measure. However, since SSTs and the accompanying experiences are expected to become a key aspect of long-term business performance (Meuter et al., 2000), it is essential to examine the customer experience, its dimensions, and the interrelations between those dimensions.

2.2.1 The interrelations between the customer experience dimensions

Since the customer experience and its dimensions play an essential role in the evaluation of a firm's performance, it is important to elaborate on how the dimensions interrelate. Krishna (2012) investigates four of the five dimensions of the customer experience in relation to each other. Krishna (2012) primarily focuses on sensory marketing – so corresponding to the sensorial dimension – and how this affects the other dimensions through perception. Chen and Lin (2018) state that the sensorial experience is essential in building an emotional connection with a product and/or brand. They state that "it is the foundation of the five modules of experiential marketing and is often the most direct and effective method in creating a lasting impression on customers" (2018, p. 553). The sensorial dimension of the customer experience is therefore where it all starts for a customer. The study of Krishna (2012) also states that the five senses – sight, touch, sound, taste, and smell – are activated by certain stimuli; this is biochemical and neurological in nature. The awareness or understanding of sensorial information is called a perception (Krishna, 2012). In line with this, Chen and Li (2018) and Rybanska et al. (2014) also address the importance of perception and state that the perception of the sensorial experience has a strong influence on the emotions of the customer. Whether emotions are stored in memory or not depends on the degree or intensity of the sensorial experience Rybanska et al. (2014). Memory is about storing and retrieving information and is considered a cognitive skill. As Krishna (2012) also points out, emotions and cognition work simultaneously and cognition is also affected by the senses (Krishna & Schwarz, 2014). The perception of the sensorial experience, the emotions evoked by this, and the cognitive process, together affect the behavior of a customer (Krishna, 2012; Chen & Li, 2018; Rybanska, 2014). In conclusion, the customer experience starts with the sensorial dimension and ends with the behavioral dimension. In this research, these two dimensions are central to measuring the customer experience.

2.3 Sensorial processing

The sensorial dimension addresses the five senses: sight, touch, sound, taste, and smell. Through these senses, a sensorial experience is created which aim is to arouse excitement, satisfaction, or differentiate between firms and products (Schmitt, 1999; Gentile et al., 2007). According to Krishna (2012), sensory marketing is about engaging the senses and this impacts perception, judgment, and behavior. Several studies argue that the sensorial dimension is the most important and valuable dimension (Gentile et al., 2007; Brakus et al., 2009). This corresponds with Krishna (2012), who states that the senses are activated before any other dimension. Additionally, Krishna and Schwarz (2014) state that the sensorial experience affects consumers' thoughts, feelings, and decisions. In psychology, the senses are seen as an important aspect of the experience of an individual and the behavior that an individual presents. According to the field of psychology, for the basic detection of the surrounding world, mental processes need to be involved with sensation. A pattern of neural messages is created and that represents the stimulus in the brain. Interpreting the incoming sensorial patterns and giving a personal meaning to this, is done by creating a perception with the mental processes (Agapito et al., 2013). Individuals receive sensorial stimuli from different sources around them. The specialized organs such as the eyes, the ears, and the skin sense these external stimuli. The information retrieved from these sensations enters the so-called mammalian cerebral cortex of the brain (Teichert & Bolz, 2018). When a certain sensation reaches this cortex, the sensation is interpreted and a perception of this sensation is made (Krishna, 2013). Since not all senses are part of the scope of this research, only the visual aspect of the senses will be further elaborated on.

2.3.1 Vision and visual processing

According to Krishna (2012), Gallego et al. (2022), and Pocock (1981), vision is the dominant sense when gaining environmental knowledge compared to the other senses. In fact, the visual system of humans collects 80% of all the sensorial data received from the environment (LMU Munich, 2019). Visual processing starts in the eyes and is then transmitted to the brain where a perception is made (Noton & Stark, 1971). The eye is made up of different components each having its own role in visual processing. For the sake of clarity, first, each relevant component in the human eye is explained with a visualization which can be found in Appendix 1, Figure 1. Then, the visual processing process is elaborated on.

The cornea is the part of the eye that covers the iris and the pupil, and is located at the front of the eye (Tessier-Lavigne, 2000). At the back of the eye, the retina is located (Tessier-Lavigne, 2000). The retina lines the inside of the eyeball and is a piece of tissue. The retina is considered part of the brain and as the most important part of the eye (Kolb, 2003). In the retina, sensorial neurons called photoreceptors are located. There are two types of photoreceptors, the rods, and the cones; both types respond to light. Rods are used for low-light vision (night vision) and cones are used for daylight, bright-colored vision (Kolb, 2013). One region of the retina, called the fovea, is important in visual processing. In this region, neurons shift to the side in order for the photoreceptors to receive a visual scene in its least distorted form. Most shifting happens in the center of the fovea, called the foveola (Tessier-Lavigne, 2000). The retina also contains ganglion cells. These cells transmit information via the optic nerve to the brain (Tessier-Lavigne, 2000). The ganglion cells have different sizes of receptive fields, which is the part of a cell that is sensitive to a certain stimulus. The optic nerve is the last relevant component of the eye. This component connects the eyes to the brain (Tessier-Lavigne, 2000; Kolb, 2013). The visual processing process consists of 4 steps. Each step will be further elaborated on afterward. The first step of the process is that light enters the eye from a visual scene and this light is projected onto the retina. The second step is that neurons in the fovea region of the retina shift to the side in order to receive the light. The third step is that the photoreceptors (cones) in the fovea respond to the light and receive the visual scene. The fourth and last step is that the ganglion cells in the retina transmit the information from the photoreceptors (cones) to the brain. This happens via the optic nerve (Tessier-Lavigne, 2000; Kolb, 2013).

The first step is that light from a visual scene enters the eye through the cornea. Next, this light is projected onto the retina, where the photoreceptors and ganglion cells are located (Tessier-Lavigne, 2000). The second step is that the neurons in the retina need to shift to the side. This happens in the fovea region of the retina, and most shifting is done in the center of the fovea, called the foveola. This process results in a projection of the visual scene onto the fovea in its least distorted form (Tessier-Lavigne, 2000). The third step is that the photoreceptors in the fovea region respond to the light and receive the visual scene in its least distorted form. The light that is received and processed in this step, is specifically daylight because the fovea region contains most of the cones (Tessier-Lavigne, 2000). After the cones have responded to this daylight and have processed the visual scene, the fourth step is that ganglion cells transmit the information. The ganglion cells in the fovea region have small

receptive fields, and therefore the visual acuity is the greatest here. This entails that the visual system is able to see the scene the sharpest in this region. The ganglion cells in the periphery of the retina have larger receptive fields, and visual acuity is low here (Tessier-Lavigne, 2000). Via optic nerve fibers, the ganglion cells transmit information from one cone in the fovea region to the visual centers in the brain (Kolb, 2013).

Where the eye is fixated on, is determined by two factors that interact complementarily and dynamically. The first one is bottom-up factors, which entail that the eye movements are influenced by a stimulus-driven perspective. It is defined as "stimulus parts that extract fixations independent of the internal state of an observer" (Schütt et al., 2019, p. 2). This entails that the external environment and the saliency of this influences where the eyes will be fixated on and which part of the visual scene is processed, regardless of the importance of this for a decision (Orquin & Mueller Loose, 2013). The second one is top-down factors, which influence eye movements from a goal-driven perspective. It is defined as "cognitive influences on the chosen fixation locations based on the current aims of an observer varying with task demands and memory" (Schütt et al., 2019, p. 2). This entails that eye movements are driven by the brain and a certain goal determines where the eyes will be fixated on (Orquin & Mueller Loose, 2013). When giving a specific task, top-down factors determine that visual attention increases toward stimuli that are relevant for that task (Orquin & Mueller Loose, 2013). These highly complex processes and their interaction with the limited capacity of the short-term memory, explain why individuals cannot process a visual scene all at once (Ballard et al., 1995). The part of a visual scene where the focus is on is processed by the fovea region in the retina. Therefore, this part of a visual scene is the sharpest. However, the visual scene consists of more than just the area of focus. This is processed by the peripheral region of the retina and thus cannot be seen sharply. Therefore, individuals are forced to process a visual scene through selective attention, which means that the most relevant stimuli for a task are filtered out (Agapito et al., 2013; Dayan et al., 2000).

When a visual scene is processed, the fovea region of the retina focuses on one area of the visual scene; this is called a fixation point. When the focus shifts to another area, another fixation point occurs and then this path from point one to point two is called a saccade. In this way, it can be investigated how an individual processes a visual scene with his/her eye movements and how someone's gaze behavior or visual attention progresses. With this understanding, it can be examined how a customer's vision reacts to a confusing situation involving chatbots.

2.4 The behavioral dimension

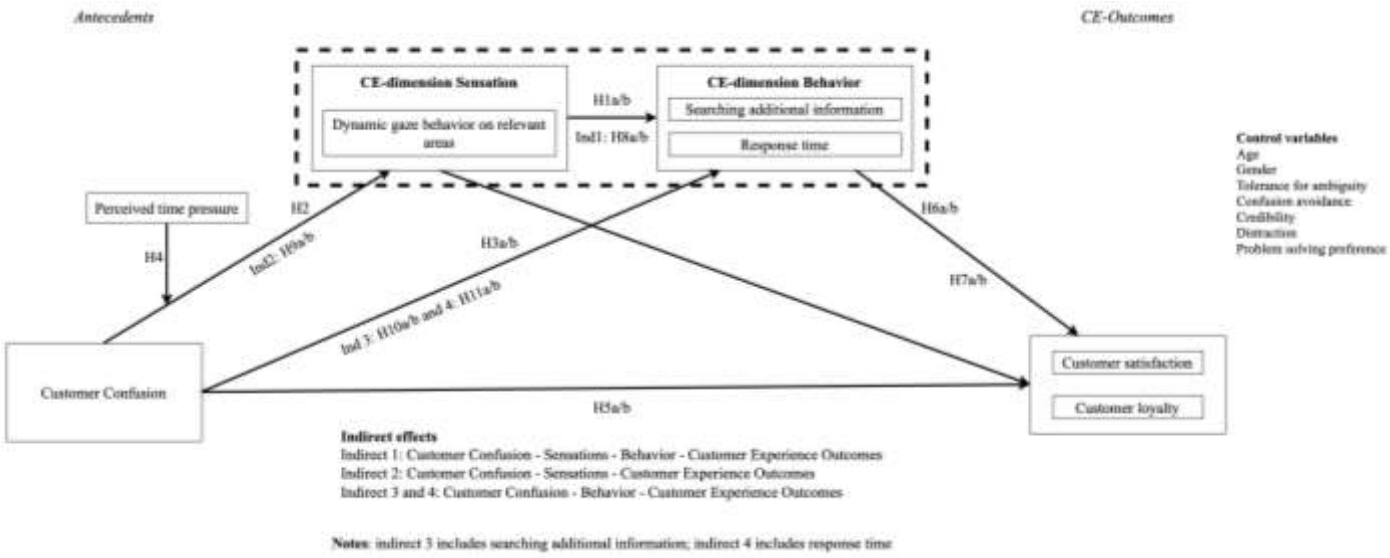
The behavioral dimension includes the effect the customer experience dimensions have on customers physically. So, the customer experience triggers a certain action, changes a certain habit, or modifies a lifestyle (Brun et al., 2017). This behavior is part of the experience itself, rather than an outcome of the customer experience (Brun et al., 2017). Schmitt (1999) views the behavioral dimension the same and states that the behavioral dimension targets physical experiences, alternative lifestyles, and interactions. Brun et al. (2017) and Pekovic and Rolland (2020) argue that the behavioral dimension is about investing time and effort in sharing information, making suggestions, engaging in decision-making processes, and participating in co-creation, which leads to engagement with the firm (Brun et al., 2017).

In this research, the behavioral dimension encompasses behavior that individuals are able to show during a human-robot interaction. In human-robot interaction, a certain action of a customer can be triggered when the customer gets confused. Customers are able to deploy several confusion-reduction strategies when the confusion level becomes uncomfortable. Mitchell and Papavassiliou (1997) mention six confusion-reduction strategies. The first one is doing nothing and the second one is postponing or abandoning the purchase. These two strategies do not reduce confusion but rather are used for coping with confusion (Drummond, 2004). The postponement or abandonment of a purchase allows customers to use one of the other four strategies to reduce confusion proposed by Mitchell and Papavassiliou (1997). One of those strategies is to involve others and to share a decision. Although Drummond (2004) argues that this strategy might not reduce the confusion immediately, it adds another dimension to a decision that might increase the customer's comfort. Another strategy is to clarify the buying goals to assess if the potential product or service actually meets the customer's needs (Drummond, 2004). This is usually done by another strategy, namely seeking additional information. This confusion-reduction strategy is one of the most common strategies to use when confused. Additionally, Kuhlthau (1991) argues that confusion is the starting point of an information search process. For this strategy to reduce confusion, the information must be of acceptable quality. If this is not the case, confusion will still be present in the form of information overload (Drummond, 2004). When this happens, the fourth and last strategy – narrowing down the set of alternatives – is an appropriate strategy to utilize according to Drummond (2004). This is done by defining criteria and choosing the best option based on these criteria (Drummond, 2004).

2.5 Hypotheses development

The relationships between customer confusion and dynamic gaze behavior, dynamic gaze behavior and searching for additional information, response time, and customer satisfaction and customer loyalty are shown in Figure 1 below. The subsequent sections provide the hypotheses. First, the interrelations between the customer experience dimensions are elaborated on. Second, the drivers of the customer experience are discussed. Third and last, the outcomes of the customer experience are addressed.

Figure 1. Conceptual model



2.5.1 Interrelation between the sensorial dimension and behavioral dimension

Based on the theories discussed in 2.2.1, it can be concluded that the dimensions of the customer experience follow a sequence. First, the sensorial dimension is affected by a specific stimulus. The cognitive and emotional dimensions perceive this stimulus and thoughts and emotions are formed. These two dimensions are beyond the scope of this research. The perception of the sensorial experience, and the thoughts and emotions evoked by this, together influence the behavioral dimension. Due to the fact that the customer experience starts with the sensorial dimension, the subsequent section elaborates on dynamic gaze behavior on relevant areas. Afterward, the behavioral dimension is discussed in more detail.

Dynamic gaze behavior on relevant areas

Gaze behavior is the eye movements of an individual (Ziv, 2016) and the concept is closely related to visual attention (Ziv, 2016; Itier & Batty, 2009). Gaze behavior typically consists of two features. First, fixation is a form of gaze behavior that entails that the individual's gaze is

maintained at one point on the screen (Lallé et al., 2016). According to Salminen et al. (2018, p. 111), "fixations capture the direction of a user's attention and therefore indicate where information acquisition and processing are possibly taking place". Fixations capture overall gaze activity on a screen when acquiring information (Lallé et al., 2016; Meißner & Oll, 2019). Second, a saccade is a form of gaze behavior that entails that an individual's eye movements quickly go from one fixation point to another. Dynamic gaze behavior is defined by Otterbring et al. (2016, p. 166) as "a frequent, widely distributed viewing pattern". This entails that customers who show more dynamic gaze behavior, have a higher number of fixations and that these fixations last longer. Fixations and their duration are appropriate measures of visual attention and thus help to understand where customers are focusing on (Otterbring et al., 2016). When confused, individuals have the urge to resolve this confusion by searching for additional information (Drummond, 2004; Kuhlthau, 1991). When this need is present, individuals fixate their eyes more and longer on informative (relevant) parts of an interface (Enders et al., 2021).

Searching for additional information

The confusion-reduction strategy of searching for additional information is one of the most common strategies to use when confused (Drummond, 2004). Additionally, Kuhlthau (1991) argues that confusion is the starting point of an information search process. In order for this strategy to reduce confusion, the content and the readability of the information must be of high quality. Additionally, the information should correspond to the target audience (Drummond, 2004). According to Brandtzaeg and Følstad (2017), chatbots can be used for many purposes, such as (additional) customer service, providing information, entertainment purposes, and social and/or emotional support. The motivations and reasons to use chatbots were empirically investigated in a study conducted by Brandtzaeg and Følstad (2017). In their study, they found that the majority of individuals use chatbots to gain information or to ask for assistance. Therefore, a chatbot is an interface to search for additional information. Due to the fact that searching for additional information is one of the most common confusion-reduction strategies and also one of the main reasons why customers use chatbots, the behavioral dimension is focused on this concept.

Response time

An important behavioral concept that is related to visual search, information search, and information processing via an interface is reaction time (Dörk et al., 2011; Deary et al., 2001;

Pachella, 1973). Reaction time (or response time) in cognitive psychology is defined as "the interval between the presentation of a stimulus to a subject and the subject's response" (Pachella, 1973, p. 6). Since information search is a coping strategy for confusion (Kuhlthau, 1991), reaction time is also related to confusion. To resolve confusion, thinking, engaging in careful deliberation, and revising existing mental model is required (D'Mello et al., 2014). Based on Daniel Kahneman's book "Thinking Fast and Slow" (2011), individuals have two systems for cognitive processing (Kannengiesser & Gero, 2019). System 1 is for fast thinking, which is intuitive, effortless, and with recognized patterns. On the contrary, system 2 is for slow thinking, which is analytic, deliberate, and requires cognitive effort (Tay et al., 2016; Kannengiesser & Gero, 2019). Based on this dual-processing theory, individuals' system 2 is activated to resolve confusion. Due to this, the reaction time of individuals is slower. Krajbich et al., (2015), Stupple et al., (2013), and Gill and Prowse (2017) all argue in line with this by stating that response times are slower when deliberate cognitive processes (utilized for resolving confusion), are activated.

The discussion above argues that an individual searches additional information via an interface when encountering confusion. During this process, the response time is slower and therefore longer. However, based on the theories discussed in 2.2.1 and 2.5.1, the sensorial dimension of the customer experience is activated first before anything else.

The effect of dynamic gaze behavior on relevant areas on searching for additional information and response time

Examining dynamic gaze behavior gives insight into how individuals search for or process information (Enders et al., 2021). When searching for or processing additional information, individuals have more and longer fixations on informative, relevant parts of an interface since useful information can be found there (Enders et al., 2021). In addition, Meißner and Oll (2019) argue that individuals only acquire information when their eyes are fixated. Based on this, it can be concluded that confusion triggers the coping strategy of searching for additional information. Before an individual can acquire information, their eyes need to be fixated on informative parts (Meißner & Oll, 2019). Therefore, visual processing is first required before an individual can act on resolving this confusion. Therefore, dynamic gaze behavior on relevant areas affects searching for additional information. Additionally, dynamic gaze behavior on relevant areas entails that individuals have more and longer fixations on informative parts of an interface. More and longer fixations on relevant areas mean that an

individual is cognitively processing and retrieving this information (Granka et al., 2008). Cognitive effort is a characteristic of system 2, activated when an individual revises mental models and engages in careful deliberation. When cognitive effort is required, response times of individuals tend to be longer (Krajbich et al., 2015; Stupple et al., 2013; Gill and Prowse, 2017). Therefore, the following hypotheses are formulated:

H1a: Customer's dynamic gaze behavior on relevant areas in a chatbot context has a positive effect on customer's search for additional information
H1b: Customer's dynamic gaze behavior on relevant areas in a chatbot context has a positive effect on customer's response time.

2.5.2 Drivers of customer experience

The eye movements (gaze behavior) of individuals are influenced by bottom-up factors (environment) and top-down factors (certain tasks) as explained in the section on visual processing. As discussed by Anninou and Foxall (2019), confusion is a response to the environment and the conditions present in that environment. Therefore, confusion is a bottom-up influence because a chatbot interaction is part of the environment. Furthermore, a concept that is important when performing a certain task is perceived time pressure (Maruping et al., 2015). Time pressure is an evaluation of a task environment and is a subjective feeling of having a lack of time when completing a task (Maruping et al., 2015; Ordóñez et al., 2015). A perception of time pressure is made within the mental processes in the brain (Krishna, 2013; Agapito et al., 2013). Due to the fact that a perception starts from within the brain, perceived time pressure starts from the person itself, rather than from the external environment. So, perceived time pressure is considered a top-down influence.

The two processes discussed above interact and this interaction strengthens attention capture and the encoding of the visual short-term memory. The visual attention that individuals devote to visual scenes is therefore determined by the interplay of bottom-up and top-down influences (Orquin & Mueller Loose, 2013). Thus, the eye movements of a customer, are determined by a combination of the state of confusion (bottom-up) and the perceived time pressure (top-down) that the customer encounters.

Confusion

Confusion can easily occur when interacting with a chatbot (Nicolescu & Tudorache, 2022). Nicolescu and Tudorache (2022) argue that customers can get confused when the messages of

the chatbot are not fully understood. Additionally, Abd-Alrazaq et al. (2021) found that individuals sometimes face difficulty in using a chatbot due to the fact that it is unclear when and how to react to it. When a chatbot makes an error and therefore confuses individuals, Cahya et al., (2019) argue and empirically show that individuals' gaze towards the screen is longer in error situations. This is in line with the research of Salminen et al. (2019), which states that a longer gaze duration is an indication of confusion when retrieving information. Additionally, Granka et al. (2008) state that the amount of time spent fixating on certain items is closely related to the degree of cognitive processing and that information processing is primarily evaluated with eye fixation metrics. Based on the eye-mind hypothesis of Just & Carpenter (1980), Wu & Liu (2022, p. 552) state that the "eyes fixate on the entity with which the mind is engaged". Drawing on this hypothesis, confused customers are likely to pay more attention to their points of confusion. Moreover, Salminen et al., (2019) state that confused individuals show more fixations than non-confused individuals. Fixations indicate where an individual devotes his/her attention to and is also related to the level of difficulty and depth of information processing. Since confusion indicates a more complex information processing process (Turnbull et al., 2000), confused customers are likely to have more fixations.

Based on the discussion above and the fact that individuals need to fixate on informative, relevant areas in order to resolve this confusion (Meißner & Oll, 2019), confused customers are likely to show more fixations and longer durations of fixations on relevant areas. Since these two aspects are part of dynamic gaze behavior, the following hypothesis is formulated:

H2: Customer confusion in a chatbot context has a positive effect on a customer's dynamic gaze behavior on relevant areas

When encountering confusion, an individual desires to resolve this confusion. One of the most common ways to do this is by searching for additional information to reduce this confusion. In order for this strategy to work, the information should be of high quality and should align with the target audience (Drummond, 2004). Individuals need to engage in careful deliberation and require to revise mental models when they desire to resolve confusion. Based on the explanation above about the dual-processing theory by Kannengiesser and Gero (2019), individuals' system 2 is activated when confusion needs to be resolved. Since system 2 is about analytic, deliberate thinking, response times of individuals are slower. Several studies correspond with this view by that reaction and/or

response times are slower when deliberate cognitive processes (utilized for resolving confusion), are activated (Krajbich et al. 2015; Stupple et al. 2013; Gill and Prowse, 2017). Based on the discussion above, the following hypotheses are formulated:

H3a: Customer confusion in a chatbot context has a positive effect on customer's search for additional information
H3b: Customer confusion in a chatbot context has a positive effect on customer's response time

Perceived time pressure

Time pressure is defined as "the difference between the amount of available time and the amount of time required to resolve a task" (Hui et al., 2020, p. 282). When performing a task, the feeling of time pressure is what determines the decision. Therefore, time pressure is an evaluation of a task environment and is a subjective feeling of having a lack of time when completing a task (Maruping et al., 2015; Ordóñez et al., 2015; Hui et al., 2020). This entails that, under the same circumstances, different people may experience different time pressures. A perception of the time is made within the mental processes in the brain (Krishna, 2013; Agapito et al., 2013). Time pressure influences the amount of information that can be processed properly by an individual (Iyer, 1989). Goeschl and Lohse (2018) argue that time pressure can therefore cause individuals to get confused since they have insufficient time to collect and interpret complete information (Pieters & Warlop, 1999). Regarding dynamic gaze behavior, Pieters and Warlop (1999) argue that individuals can cope with time pressure by using the acceleration strategy or the filtration strategy. The acceleration strategy entails that individuals speed up their visual information collection and processing. In terms of gaze behavior, the combination of confusion and time pressure means that the duration of fixations on the relevant parts is reduced (Pieters & Warlop, 1999). The filtration strategy occurs when confused individuals skip certain elements of information and thus fixate less. Based on the discussion above, dynamic gaze behavior is affected by the interplay between confusion (bottom-up) and time pressure (top-down). When a customer is confused and perceives no time pressure, more dynamic gaze behavior is apparent. However, when a customer is confused and perceives time pressure, then less dynamic gaze behavior on relevant areas is apparent (due to fewer fixations and shorter fixations). Therefore, the following hypothesis is formulated:

H4: Customer's perceived time pressure in a chatbot context moderates the relationship between customer confusion and customer's dynamic gaze behavior on relevant areas such that a) the effect is strengthened for low levels of perceived time pressure, and b) the effect is attenuated for high levels of perceived time pressure

2.5.3 Outcomes of the customer experience

Customer satisfaction and customer loyalty

Customer satisfaction is defined by Zaid and Patwayati (2021, p. 985) as "a feeling that appears on the results of a purchase evaluation which is a comparison between actual perception and expectation". Customer loyalty is defined as "the intention to perform a diverse set of behaviors that indicate the motivation to maintain a relationship with a brand or company" (Torabi et al., 2021, p. 685). When a customer is loyal to a product, this customer more likely shares good experiences with others (Akbari et al., 2016). Customer satisfaction and customer loyalty are both outcomes of customer confusion and the customer experience.

Customer confusion

Due to the fact that customer confusion entails that customers have difficulty with selecting, interpreting, and evaluating facets of a product or service, customers' ability to choose the option that is most beneficial to them is reduced. As a result of this, customers' satisfaction decreases (Edward & Sahadev, 2012). Moreover, as Moon et al. (2016) point out, especially when individuals get confused due to similarities or an overload of information, products, or offerings, satisfaction decreases. This is because individuals are not able to discriminate between the similarities or the excessive number of offerings and because confused customers need to invest more time, effort, and energy to process information.

Edward and Sahadev (2012) argue that confusion reduces customer loyalty toward a product and/or a service. When confusion increases, customers experience a lack of trust and a lack of confidence in the product, the service, of the firm. Therefore, their loyalty decreases. This is in line with the research of Reichheld et al. (2000), which argues that confusion harms loyalty, due to a lack of trust. Additionally, Mitchell and Papavassiliou (1999) also argue that a confused customer is less likely to be a loyal customer due to the fact that the customer was not able to choose the most beneficial option available to them.

Chatbots are primarily designed to enhance customer satisfaction and enhance customer loyalty, however, confusion can easily occur while interacting with a chatbot (Nicolescu & Tudorache, 2022). The design of the message is especially important because

when this is not properly executed, the user might not understand it and will get confused (Waghmare, 2019; Nicolescu & Tudorache, 2022). Individuals sometimes face difficulty when using a chatbot because it is not clear how to react to it (Abd-Alrazaq et al., 2021). Specifically, it is hypothesized that customer confusion caused by a chatbot has a negative effect on customer satisfaction and customer loyalty.

H5a: Customer confusion in a chatbot context has a negative effect on customer satisfaction.
H5b: Customer confusion in a chatbot context has a negative effect on customer loyalty

Customer experience

Numerous studies argue that the customer experience has an impact on key outcomes such as customer satisfaction and customer loyalty (Lemon & Verhoef, 2016; Klaus & Maklan, 2013; Kusumawati & Rahayu, 2020; Zaid & Patwayati, 2021; Mokha & Kumar, 2022). The customer experience is an important indicator of success when building relationships with customers. The interactions between the firm and the customer during an experience are able to make an impression on customers and can influence the way a customer assesses a product or service. When the customer experience is evaluated as positive or memorable, then this will have a positive effect on customer satisfaction and loyalty (Zaid & Patwayati, 2021; Mohka and Kumar, 2021; Klaus and Maklan, 2013). Regarding loyalty, the research of Kusumawati & Rahayu, (2020) proves that the evaluated quality of the experience is a determinant of customer loyalty. Regarding satisfaction, the positive effect of customer experience quality on customer satisfaction is in line with the equity theory. This theory states that customer satisfaction increases when the experience is evaluated best when the sacrifice in terms of time, money, and effort is minimal and when the perceived value received from a product or service exceeds expectations (Kusumawati & Rahayu, 2020). Mohka and Kumar (2022) also state that an experience is seen as favorable when it requires minimal effort and time. Searching for additional information during an experience could be seen as an extra effort for the customer, who also has to devote more time to the experience. Additionally, a longer response time results in a longer experience with more effort. Therefore, the behavioral dimension of searching for additional information has a negative effect on customer satisfaction and customer loyalty.

H6a: Customer's search for additional information in a chatbot context has a negative effect on customer satisfaction

H6b: Customer's search for additional information in a chatbot context has a negative effect on customer loyalty

H7a: Customer's response time in a chatbot context has a negative effect on customer satisfaction
H7b: Customer's response time in a chatbot context has a negative effect on customer loyalty

2.5.4 Serial mediation and mediation

The before-mentioned hypotheses provide insight into the relationship between customer confusion and customer experience outcomes (satisfaction and loyalty), mediated by the sensorial dimension (dynamic gaze behavior) and the behavioral dimension (searching for additional information) of the customer experience.

H8a: Customer's dynamic gaze behavior on relevant areas and customer's search for additional information serially mediate the relationship between customer confusion and customer satisfaction in a chatbot context, such that (a) customer confusion has a positive effect on customer's dynamic gaze behavior on relevant areas, (b) customer's dynamic gaze behavior on relevant areas has a positive effect on customer's search for additional information, and (c) customer's search for additional information has a negative effect on customer satisfaction

H8b: Customer's dynamic gaze behavior on relevant areas and customer's search for additional information serially mediate the relationship between customer confusion and customer loyalty in a chatbot context, such that (a) customer confusion has a positive effect on customer's dynamic gaze behavior on relevant areas, (b) customer's dynamic gaze behavior on relevant areas has a positive effect on customer's search for additional information, and (c) customer's search for additional information has a negative effect on customer loyalty

Besides this serial mediation effect, two mediation paths became clear. As mentioned in hypothesis 2, confusion affects the dynamic gaze behavior of a customer (Salminen et al., 2019). In addition, confused customers are likely to show certain behaviors to reduce this confusion. One of the most common is searching for additional information (Drummond, 2004). These customer experience dimensions influence customer satisfaction as well as

customer loyalty (Klaus & Maklan, 2013; Kusumawati & Rahayu, 2020; Zaid & Patwayati, 2021; Mokha & Kumar, 2022; Eskiler & Safak, 2022).

H9a: Customer's dynamic gaze behavior on relevant areas mediates the relationship between customer confusion and customer satisfaction in a chatbot context, such that a) customer confusion has a positive effect on customer's dynamic gaze behavior on relevant areas, and b) customer's dynamic gaze behavior on relevant areas has a negative effect on customer satisfaction
H9b: Customer's dynamic gaze behavior on relevant areas mediates the relationship between customer confusion and customer satisfaction in a chatbot context, such that a) customer confusion has a positive effect on customer's dynamic gaze behavior on relevant areas, and b) customer's dynamic gaze behavior on relevant areas has a negative effect on customer loyalty

H10a: Customer's search for additional information mediates the relationship between customer confusion and customer satisfaction in a chatbot context, such that a) customer confusion has a positive effect on customer's search for additional information, and b) customer's search for additional information has a negative effect on customer satisfaction
H10b: Customer's search for additional information mediates the relationship between customer confusion and customer satisfaction in a chatbot context, such that a) customer confusion has a positive effect on customer's search for additional information, and b) customer's search for additional information has a negative effect on customer loyalty

H11a: Customer's response time mediates the relationship between customer confusion and customer satisfaction in a chatbot context, such that a) customer confusion has a positive effect on customer's response time, and b) customer's response time has a negative effect on customer satisfaction
H11b: Customer's response time mediates the relationship between customer confusion and customer satisfaction in a chatbot context, such that a) customer confusion has a positive effect on customer's response time, and b) customer's response time has a negative effect on customer loyalty

3. Methodology

The conceptual model provided in Figure 1 is tested with a lab experiment in a chatbot context. The experiment is set up in a way that allows researchers or academics to enhance the understanding of how people experience confusion with chatbots in any industry.

3.1 Design

The experiment is this research in set up using a Wizard of Oz approach. The Wizard of Oz experiment is a method that is widely used in the field of human-robot interaction (Steinfeld et al., 2009). In these experiments, the participants think that they are interacting with the technology itself, while it is the researcher who controls and performs the task of the technology (Steinfeld et al., 2009). Thus, in this research, participants feel that they are interacting with a real chatbot, but this chatbot is controlled and managed by one of the researchers. The Wizard of Oz experiment method is a well-established and accepted experiment design to evaluate human behavior (Steinfeld et al., 2009).

The experiment in this research uses a 2 (confusion: yes vs. no) x 2 (perceived time pressure: high vs. low), between-subjects experimental design. Two factors are manipulated, namely confusion and perceived time pressure. In total, there are four conditions where participants are randomly assigned to. Condition one is where participants encounter a situation with no confusion and no time pressure. Condition two is a situation where participants encounter confusion, but no time pressure. In condition three, participants end up in a situation with no confusion, but with time pressure. Condition four entails that participants encounter a situation in which there is confusion and time pressure. A pre-test and manipulation check of the lab experiment is done in order to ensure that the confusion and time pressure are properly manipulated. This will be further elaborated on in a subsequent section. Participants are not informed about the actual goal of the experiment and also do not receive the full details about the experiment. This is to make sure that the participants believe they are interacting with a real chatbot instead of a human controlling the chatbot. In addition, fully informing participants will lead to a more biased data collection.

Scenario for manipulation

A manipulation scenario as credible as possible was created and designed. In this research, the focus is on a specific kind of chatbot, namely the shopping bot. Shopping bots are a new kind of chatbot that support customers with searching and purchasing products while e-shopping

(Rowley, 2000). Participants are informed that they are going to help a shopping bot (named Cody) with its design and operation by making a purchase. During this shopping experience, an error appears and the participant is unable to complete the purchase. To solve the error, the shopping bot asks for an error number. In the condition where there is no confusion, the error number is prominently shown on the error page. In the condition where there is confusion, participants see an error page with several numbers, except the error number. The time pressure manipulation is done by adding a timer of three minutes on the error page in the condition where participants need to experience time pressure. In the condition where no time pressure is present, the timer is not shown. Eventually, all participants are able to fix the error with the help of the shopping bot to ensure every participant has the same outcome. Appendix 2 contains the screenshots of the four conditions, as well as the other web pages that are included in the experiment. The questions of the manipulation check are presented in Appendix 3. The results of the manipulation check are reported in Chapter 4.

3.2 Measurement

In this experiment, two factors were manipulated: customer confusion and perceived time pressure. Therefore, no measures are needed for these two constructs. The other constructs such as dynamic gaze behavior, customer satisfaction, and customer loyalty are measured using existing scales from literature. All the constructs and their corresponding measurements are shown in Appendix 4, Table 1. A full script of the constructs and corresponding items can be found in Appendix 5.

3.2.1 Measurement of dynamic gaze behavior

The first construct that is measured is dynamic gaze behavior. Eye-tracking is a well-established and useful technique to measure the gaze behavior of participants and is used in several studies (Salminen et al., 2018; Pachman et al., 2016; Lallé et al., 2016). Eye-tracking technology has offered new insights into aspects of selective attention in recent years. With an eye-tracking device, it is possible to measure visual selective attention in real-time, in a direct and continuous way. Without a response from a participant, an eye-tracking device can precisely register the location of eye movements and thus gaze behavior. This can be done over a specific period of time (Reichenberger et al., 2020).

The measures for dynamic gaze behavior in this research are derived from Meißner and Oll (2019) and include the number of fixations and the average duration of fixations. The

number of fixations entails how many ‘focus points’ a participant records while performing the experiment. The average duration of fixations means how long one fixation lasts. Usually, a fixation lasts between 100 and 400 milliseconds (Salvucci & Goldberg, 2000). A fixation means that the customers’ gaze is maintained at one point on the screen (Salminen et al., 2018). The metrics that are used, are focused on relevant areas of the screen. Relevant areas of the screen are the parts of the screen where the information is located or where the participants can ask for information. In the experiment, two Areas of Interest (AOIs) are determined; the purchasing webpages including the error, and the shopping bot conversation. These two AOIs are the main focus of this study since these are the relevant, informative parts. However, non-relevant parts, such as the white header on the webpage, are also taken into account as an additional analysis to specifically compare between areas.

Pupil Labs Core

After conducting tests, it was decided to use the Pupil Labs Core. The Pupil Labs Core is a wearable eye-tracker and it is able to map the gaze onto a surface (Faraji et al., 2022). Due to the fact that this eye-tracker is wearable, the participant is not restricted in his/her head movements and can move and look freely at the environment he/she is in. The Pupil Labs Core can measure up to 200 Hz per eye and up to 120 Hz per eye when the resolution is higher. The eye-tracker has three cameras. The first one is a one-world camera with a fisheye field of view. This camera enables the researchers to record the environment from the point of view of the participant. The other two cameras are focused on one eye each. In this way, the Pupil Labs Core is facilitating remote eye tracking on a wide screen (Faraji et al., 2022). The Pupil Labs Core has its own software (Pupil Player) in which the AOIs can be determined and where fixations and the durations of these can be examined (Pupil Labs, 2023). For the fixations, a range between 100ms and 400ms is chosen, since this is the acceptable range according to Salvucci & Goldberg (2000).

3.2.2 Measurement of other constructs

The behavioral dimension is measured using real-time measures during the experiment, such as the number of times a participant asked for help or clarification. Additionally, one open question during the shopping bot conversation is analyzed qualitatively. Therefore, a mixed-methods approach is utilized in this research. Customer satisfaction is measured with one item derived and adapted from Chung et al. (2020). The item is measured on a seven-point Likert scale ranging from 1 (strongly dissatisfied) to 7 (strongly satisfied). Customer loyalty is

measured using two items. One item is “How likely is it that you would use this shopping experience again?” and is measured on a seven-point Likert scale ranging from 1 (highly unlikely) to 7 (strongly likely). The other item was derived from Reichheld (2003) and is also known as the Net Promoter Score (NPS). The scale for this item is adapted to correspond with the other scales used in this research. The measurement of loyalty is more focused on the shopping experience with the shopping bot instead of loyalty towards the firm that provides the product or service.

Control variables

The first control variable is tolerance for ambiguity. Individuals who are more tolerant of ambiguity are better able to process unfamiliar, complex, or dynamically uncertain stimuli or stimuli that are subject to multiple conflicting interpretations (McLain, 1993). Therefore, more ambiguate tolerant individuals are less likely to become confused because they are able to process the ambiguity better compared to individuals who show less tolerance for ambiguity. On the contrary, individuals with a low tolerance for ambiguity pause or terminate their information processing activities and/or show resistance to new information (Hoque et al., 2021). The items measuring tolerance for ambiguity are derived from McLain (1993). All items for tolerance for ambiguity are measured on a seven-point Likert scale ranging from 1 (strongly disagree) to 7 (strongly agree). The second control variable is confusion avoidance. Confusion avoidance is a personality trait that determines whether or not a customer withdraws from a confusing situation (Hoque et al., 2021). When a customer wants to avoid confusion, he/she is willing to do everything in order to make sure no confusion is present. Therefore, customers with a high level of confusion avoidance are less sensitive to confusion since they are capable of handling this confusion and eventually are able to avoid it (Hoque et al., 2021). The items measuring confusion avoidance are derived from Schweizer, Kotouc, and Wagner (2006). All items are measured on a seven-point Likert scale ranging from 1 (strongly disagree) to 7 (strongly agree). Additionally, some alternative explainability and control variables are added, such as the credibility of the shopping bot, the distraction during the experiment, and the problem-solving preference (human vs. robot). Finally, the demographic control variables age and gender are included. It is expected that older age groups encounter more complexity and difficulty while interacting with a shopping bot. Therefore, older age groups are more likely to be confused when interacting with a shopping bot, compared to younger age groups (Turnbull et al., 2000).

3.3 Procedure

As mentioned before, this experiment is set up using the Wizard of Oz experimental method First, a consent form (see Appendix 6) is handed over to the participant. This is to ensure that, among other things, the participant is giving permission to the researchers to use the data. Then, the participant will be randomly assigned to one of the four conditions, based on a number from 0-200. After this, the participant enters the lab where the experiment is performed and is instructed to put his/her phone on silent mode. The instructions given next include that the participant is going to talk to a shopping bot in English, so the participant is also informed to answer in English. In addition, the researchers tell the participant that the shopping bot will provide further instructions and that he/she should solve any issues with the shopping bot. After this, information about the equipment is provided. The piece of equipment that the participant gets is the eye-tracker, the Pupil Labs Core. The participant is able to put on the eye tracker him/herself to ensure that it is comfortable and that the eye-tracker does not hinder the experience during the experiment. Next, the eye-tracker is calibrated with the help of a laptop by adjusting the positions of the cameras if needed. To calibrate in a proper way, the participant is instructed to only move the eyes during calibration, keep head movements as minimal as possible, and keep his/her eyes as wide open as possible. After calibration is done, it is checked whether it was successful or not. The participant is instructed not to touch the eye-tracker during the experiment and to stay seated until the experiment is finished. Furthermore, it is mentioned that the participant needs to enter his/her participant ID to start the conversation with the shopping bot. This is a sign for the researchers to start the Wizard of Oz experiment. From now on, the data from the eye-tracker and the conversation with the shopping bot will be gathered. At the end, researchers enter the room again to remove the eye-tracker. A quick debrief with the participant and thanking the participant for his/her participation is the last stage of the whole experiment.

3.4 Data analysis procedure

The data analysis software programs that are used to analyze the data of the lab experiment are IBM SPSS Statistics, Version 28 (IBM Corp, 2021) and Smart PLS (Ringle, 2022). Smart PLS software can be used for variance-based structural equation modeling using the Partial Least Squares method (PLS). In structural equation modeling, two different models are tested: the measurement model and the structural model. In addition, with PLS, relationships can be

tested simultaneously and this method is a combination of factor analysis and multiple regression (Dash & Paul, 2021).

3.5 Participants

The Radboud University campus in The Netherlands is where individuals were asked to participate in the experiment. The sampling technique that is used is non-probability sampling because the participants are gathered at Radboud University and not every single person in the population has an equal chance to participate. Within the category of non-probability sampling, the method of convenience sampling is used. Convenience sampling is the appropriate method here because the lab experiment takes place at Radboud University, and the students on campus are easily accessible for participation and are available at the moment the lab experiment is performed (Emerson, 2015). A total of 200 respondents participated in the study, which entails that the criterion of the central limit theorem was reached (Anderson, 2010). Participants had an average age of 25 years and 57,8% of the participants were female.

3.6 Research ethics

The research ethics in this research are based on the Netherlands Code of Conduct for Research Integrity (2018), which is supported by the KNAW, NOW, Netherlands Association of Universities of Applied Sciences, NFU, TO2 Federation, and the VSNU. The consent form that is handed out to the participants before the experiment starts, states that participants will give permission to the researchers to use the data that is collected from them. In addition, names are not used before, during, or after the experiment to ensure that data cannot be traced back to a specific participant. The participants only get a participant ID and will not be asked to fill in personal data except for their age and name (on the consent form). The consent form is in no way traceable to any participant number. So, data is collected anonymously. The data that is collected during the experiment will not be shared with others except with the researchers that are also involved in the experiment. Data is handled confidentially since the data is only used for answering the research question and not for any secondary purposes. Furthermore, participants are able to withdraw from the experiment at any time. At the end of the experiment, participants are able to write down their email addresses in case they are interested to hear about the results of the research. This email address will in no way be linked to a participant ID.

4. Research results

To analyze the data of the experiment, IBM SPSS Statistics, Version 28 and SmartPLS (Ringle, 2022) were used. SPSS was used to check the descriptive statistics of the data and to reverse code some items. Next, SmartPLS 4 was used to validate and assess the measurement model and structural model by examining the reliability and validity statistics of these models. In addition, the structural model was assessed to check the hypotheses that were formulated.

4.1 Data preparation

4.1.1 Cleaning the data and manipulation check

Before conducting the data analysis in Smart PLS, the data first needed to be cleaned. The cleaning of the data happened in Excel where unusable data was reported as missing data. An example of unusable data is when a participant answered with an '8' while the measurement scale had a maximum of a '7'. In addition, the eye-tracking data of the two areas of interest were aggregated to better compare it with the data on the non-relevant areas. After this, a dummy variable was created for the manipulation of confusion and perceived time pressure. Next, four items in total were reversed coded in SPSS to make sure each statement was either positively or negatively formulated.

Manipulation check pre-test

To check whether the confusion and time pressure manipulation worked, a pre-test is performed. Before performing this, a confirmatory factor analysis is conducted in SPSS to see if the items of confusion and time pressure are reliable indicators. Appendix 7A shows this factors analysis. Both KMO and Bartlett met the criterion of >.50 and significance respectively (Field, 2017). Also, the communalities and factor loadings were satisfactory (<.50) (Field, 2017) after deleting one item with confusion. Items loading highly on the component were examined for reliability with Cronbach's alpha. Cronbach's alpha for confusion (.864) and for time pressure (.811) were both above the threshold of .70 (Field, 2017). Then, the manipulation check was performed by means of an independent sample t-test (see Appendix 7B). There was a statistically significant difference between confusion (M = 5.13 and SD = 1.04) and no confusion (M = 2.63 and SD = 0.95), $p < .001$. For time pressure, there was a statistically significant difference between time pressure (M = 4.87 and SD = 1.96) and no time pressure (M = 2.97 and SD = 1.06), $p = .009$). Therefore, the manipulation of both confusion and time pressure was successful. Additionally, there was no statistically

significant difference between confusion (*M* = 5.90 and *SD* = .74) and no confusion (*M* = 5.60 and *SD* = 1.27), *p* = .527 for the alternative explanation of the realness of the chatbot and the information that it provided (confusion: *M* = 4.80 and *SD* = 1.48; non-confusion: *M* = 4.90 and *SD* = 2.18), *p* = .906. Finally, there was no statistically significant difference between time pressure (*M* = 5.80 and *SD* = .92) and no time pressure (*M* = 5.70 and *SD* = 1.16), *p* = .833 for the alternative explanation of the realness of the chatbot and the information that it provided (time pressure: *M* = 5.30 and *SD* = 1.64; no time pressure: *M* = 4.40 and *SD* = 1.96), *p* = .279. The full results are presented in Appendix 7B.

Manipulation check experiment

In order to check whether the manipulation of confusion and time pressure worked properly, a manipulation check of the actual experiment is performed. Before conducting the manipulation check, a confirmatory factor analysis is performed in SPSS to see if confusion and time pressure were reliable constructs with the items used. Appendix 8A presents this factor analysis. KMO (.680) and Bartlett's test (<.001) met the criterion of >.50 and significance respectively (Field, 2017). Eventually, after deleting one item for confusion and one item for time pressure, communalities and factor loadings were satisfactory above the threshold >.20 and >.50 respectively (Field, 2017). The items loading highly on the components were examined on reliability by means of Cronbach's alpha. Cronbach's alpha for confusion was satisfactory .743 > .70 and Cronbach's alpha for time pressure was satisfactory as well .769 > .70. Next, the manipulation check was conducted by means of an independent sample t-test using SPSS (see Appendix 8B). There was a statistically significant difference between confusion (*M* = 5.31 and *SD* = .94) and no confusion (*M* = 3.34 and *SD* = 1.48), *p* < .001. For time pressure, there was a statistically significant effect between time pressure (*M* = 4.66 and *SD* = 1.80) and no time pressure (*M* = 3.63 and *SD* = 1.60), *p* < .001. It can be concluded that the manipulation of both confusion and perceived time pressure was successful during the experiment. Additionally, there was a statistically insignificant difference between confusion (*M* = 5.74 and *SD* = 1.22) and no confusion (*M* = 5.77 and *SD* = 1.22), *p* = .884 for credibility. Furthermore, there was a statistically insignificant difference between confusion (*M* = 2.35 and *SD* = 1.59) and no confusion (*M* = 2.67 and *SD* = 1.72), *p* = .173 for distraction. Finally, there is a statistically significant difference between confusion (*M* = 4.52 and *SD* = 1.76) and no confusion (*M* = 3.44 and *SD* = 1.67), *p* < .001 for problem-solving preference. This entails that participants in the confusion conditions rather would have talked to a human instead of a chatbot and got confused due to that. For time pressure,

there was a statistically insignificant difference between time pressure (*M* = 5.80 and *SD* = 1.33) and no time pressure (*M* = 5.71 and *SD* = 1.09), *p* = .621 for credibility. Additionally, there was a statistically insignificant difference between time pressure (*M* = 2.72 and *SD* = 1.78) and no time pressure (*M* = 2.30 and *SD* = 1.51), *p* = .07 for distraction. Finally, there was a statistically insignificant difference between time pressure (*M* = 3.88 and *SD* = 1.79) and no time pressure (*M* = 4.09 and *SD* = 1.81), *p* = .408 for problem-solving preference. This entails that none of the participants felt time pressure due to any other explanation. The full results of the manipulation check are presented in Appendix 8B.

4.1.2 Missing data analysis

A missing data analysis was performed using SPSS. In total 200 participants finished the experiment, however, some of these participants had some missing data. The percentages of missing data per item were all 10% or lower (see Appendix 9A, Table 9A.1), which means that the missing data is acceptable (Lee & Huber, 2021). However, when looking at the type of missing data, Little's MCAR test is significant at an alpha level of .05 (see Appendix 9A, Table 9A.2). This entails that the missing data are not Missing Completely at Random (Sainani, 2015). When looking into the data, it became clear that the variable 'age' was causing this significant result of Little's MCAR test. Since 'age' only had one missing value, the percentage of missing values is 0,5%. For this small percentage, the impact of a single imputation method is negligible and therefore justified (Sainani, 2021; Murdoch et al., 2018). The variable 'age' has a skewed distribution (see Table 9A.3) and therefore median replacement is the preferred and more robust single imputation method (Gromski et al., 2014). This replacement is done with SPSS. After this, Little's MCAR test was not significant at an alpha level of .05 (see Table 9A.2) and, therefore the missing data can be interpreted as Missing Completely at Random.

4.2 Evaluation of the measurement model

The evaluation of the measurement model is done within SmartPLS and is done by conducting a confirmatory factor analysis (CFA) and by examining statistics for factor loadings (internal reliability), construct reliability, convergent validity, and discriminant validity. The threshold for factor loadings is that these loadings should be 0.5 or higher, but an ideal factor loading is 0.70 or higher (Hair et al., 2019, p. 663). Moving on, construct reliability values should be above 0.70 (Hair et al., 2019, p. 663) and are only given in

SmartPLS for constructs containing more than one item. Furthermore, convergent validity is measured with the Average Variance Explained (AVE) and this value should be above 0.50 (Hair et al., 2019, p. 663). Lastly, discriminant validity is examined with the Heterotrait-Monotrait (HTMT) ratio of correlations. The HTMT values should be below 0.85 (Hair et al., 2019, p. 776).

The initial model (see Figure 9B.1 and 9B.2, Appendix 9B) did not meet these criteria at first (see Table 9B.1 and 9B.2 in Appendix 9B), due to negative and low loadings on some constructs. A negative loading indicates that an item measures different dimensions of the construct and therefore does not measure the construct in the correct way (Asnawi et al., 2012). Therefore, the two constructs that measured dynamic gaze behavior were split up into four constructs with one item. In addition, the two constructs measuring the CE-dimension of behavior also did not meet the criteria (see Table 9B.3 and 9B.4 in Appendix 9B). Therefore, some items that loaded negatively and/or low were deleted. The following items were sequentially deleted: the number of clicks on the information button and response time error page. Furthermore, the control variables 'confusion avoidance' and 'tolerance for ambiguity' did not meet the criteria of construct reliability and convergent validity (see Table 9B.3 and 9B.4 in Appendix 9B). In order to improve these statistics, the following items were sequentially deleted: ambiguity Q2, confusion avoidance Q1, and confusion avoidance Q3.

Subsequently, acceptable statistics for both reliability and validity were achieved (see Table 9B.5 – 9B.7 and Figure 9B.3 and Figure 9B.4 in Appendix 9B). The factor loadings are all above the ideal threshold of 0.70, except ambiguity Q3. This item has a loading of 0.604 but is not deleted from the analysis since the construct reliability and convergent validity of the construct 'tolerance for ambiguity' are acceptable and the loading is above the minimum of 0.50. Additionally, the construct reliability is only given for Loyalty (.823) and Tolerance for ambiguity (.762). The value of the composite reliability Rho C is used because this does not assume that each item has an equal contribution to the construct (Achjari, 2015). The convergent validity is also only given for Loyalty (.701) and Tolerance for ambiguity (.626) and is above the threshold of .50. Finally, discriminate validity was achieved because all the HTMT values did not exceed the threshold of 0.85. A summary of the constructs with the corresponding composite reliability values, convergent validity values, and factor loadings is shown in Appendix 9B, Table 9B.8.

The last statistic is the model fit. The model fit is examined by looking at the SRMR statistic. According to Ringle (2022), it is still unclear in practice if the value of the saturated model or estimated model should be looked at. Here, the saturated model fit is 0.043, which is

below the desirable threshold of 0.08. However, the estimated model fit is 0.130, which is above the desirable threshold of 0.08. Since it is not clear which one to look at, both are taken into consideration and it can be concluded that the model has a moderate fit. Additionally, the Goodness-of-Fit (GoF) Index was utilized to examine the model fit. Wetzels et al. (2009) state that 0,1 is the small threshold, 0,25 is the medium threshold, and 0,36 is the large threshold. Appendix 9B, Table 9B.9 presents the calculation of the GoF Index. The GoF Index for this model is 0,37, which exceeds the threshold for a large model fit.

4.3 Evaluation of the structural model

4.3.1 Collinearity and Coefficient of Determination

Examining the collinearity among predictor constructs is necessary because otherwise, the path coefficients might be biased (Hair et al.,2019, p. 790). This is done by looking at the Variance Inflation Factor (VIF) values. The VIF values should be below the threshold of < 3.0. As can be seen in Appendix 9C, Table 9C.1, the VIF values are all < 3.0 and, therefore no issues with collinearity are present and this criterion is met. Second, the coefficient of determination (R^2) is examined, which is a measure of the predictive power of the model. The higher the value of R^2, the higher the explanatory power of the model and thus the better the prediction of the dependent variable (Hair et al., 2019, p. 260). However, in this research, the value for the adjusted R^2 is examined. This value takes the sample size and the complexity of the model into account (Hair et al., 2019, p. 260) and therefore is the recommended statistic (Heinzl & Mittlböck (2002). An adjusted R^2 of 0.01 is weak, 0.09 is moderate, and 0.25 is strong (Jacobs & Korzilius, 2022). Based on these thresholds, the constructs Average duration non-relevant areas (0.014), Average duration relevant areas (0.017), and Total fixations relevant areas (0.033) have weak predictive power. The construct Total fixations non-relevant areas has a negative adjusted R^2. According to Heinzl and Mittlböck (2002), a negative adjusted R^2 indicates that there is no explained variance at all. Therefore, this construct is deleted from the analysis. Moreover, the construct Searching additional information (0.170) has moderate predictive power. Finally, the constructs Loyalty (0.231), Response time (0.279), and Satisfaction (0.263) have strong predictive power.

4.3.2 Effect sizes

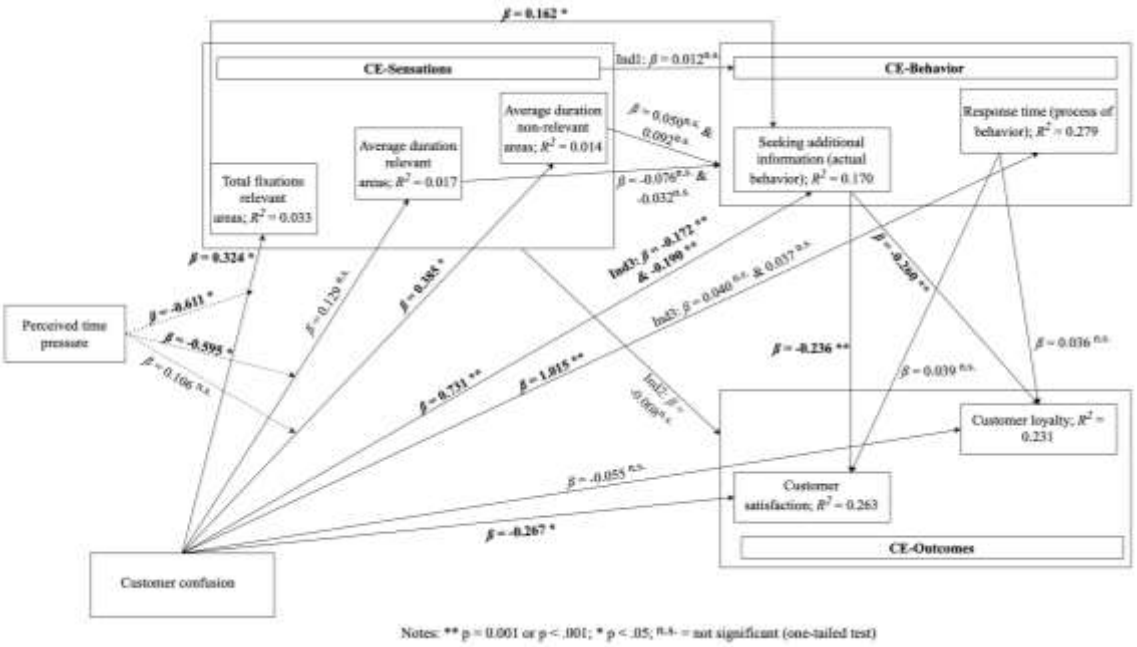
The effect size is defined by Hair et al. (2019, p. 780) as “The effect size represents the change in de R^2 value when a specified exogenous construct is omitted from the model”. The

appropriate measure for this is Cohen f^2. Effect sizes of 0.02, 0.15, and 0.35 indicate a weak, moderate, and strong effect, respectively. An effect size smaller than 0.02 indicates that there is no effect (Hair et al., 2019, p. 780). The effect sizes range from 0.000 to 0.338. The smallest effect size is from Total fixations relevant areas to Satisfaction (0.000). The largest effect size is from Confusion to Response time (0.338). All the effect sizes are shown in Appendix 9C, Table 9C.4.

4.3.3 Path coefficients

Path coefficients are the values that represent the relationship between the constructs in the model (Hair et al., 2019, p. 762). In SmartPLS, these path coefficients are standardized to make comparison easier (Hair et al., 2019, p. 208). Therefore, a higher path coefficient (ßeta coefficient) indicates a stronger relationship than a lower path coefficient, when significant. The results of the structural model evaluation are shown in Figure 2 below.

Figure 2. Results structural model



To begin with, the results indicate that there is a statistically significant, positive effect of total fixations relevant areas on searching for additional information (*ß* = 0.162; *p* = .028; R^2 = 0.170). Surprisingly, there is a statistically insignificant effect of average duration relevant areas on searching for additional information (*ß* = -0.076; *p* = .154; R^2 = 0.170). Therefore, H1a is partially supported. Additionally, there is a statistically insignificant effect of total fixations relevant areas on response time (*ß* = 0.087; *p* = .095; R^2 = 0.279) and of average

duration relevant areas on response time (ß = -0.032; p = .350; R^2 = 0.279). Therefore, H1b is not supported. Moving on, a dummy variable was created for confusion and the dummy 'YES' confusion was included in the analysis. There was a statistically significant, positive effect of customer confusion on total fixations relevant areas (ß = 0.324; p = .043; R^2 = 0.043). A statistically insignificant effect was shown of customer confusion on average duration relevant areas (ß = 0.129; p = .248; R^2 = 0.017). Therefore, H2 is partially supported. Furthermore, there was a statistically significant, positive effect of customer confusion on searching for additional information (ß = 0.731; p < .001; R^2 = 0.170) and on response time (ß = 1.015; p < .001; R^2 = 0.170). Therefore, H3a and H3b are supported.

Moreover, a dummy variable for perceived time pressure was included in the analysis. The dummy variable "YES" was included, meaning that a score of 0 meant 'no perceived time pressure' and a score of 1 meant 'perceived time pressure'. There was a statistically significant, negative moderation effect of perceived time pressure on the relationship between customer confusion and average duration relevant areas (ß = -0.595; p = .018; R^2 = 0.017). The Simple Slope analysis is shown in Appendix 9C, Figure 9C.1 In a situation of confusion and time pressure, the average duration on relevant areas is lower compared to a confusing situation with no time pressure. This entails that, in a situation of confusion and perceived time pressure, customers have shorter fixations on relevant areas compared to when they do not perceive time pressure in a confusing situation. There was a statistically insignificant moderation effect on the relationship between customer confusion and total fixations relevant areas (ß = 0.106; p = .354; R^2 = 0.033). Therefore, H4 is partially supported. To continue, there was a statistically significant, negative effect of customer confusion on customer satisfaction (ß = -0.267; p = .044; R^2 = 0.263) and a statistically insignificant effect of customer confusion on loyalty (ß = -0.055; p = .379; R^2 = 0.231). Based on this, H5a is supported but H5b is not supported. Furthermore, there is a statistically significant, negative effect of searching for additional information on customer satisfaction (ß = -0.236; p < .001; R^2 = 0.263) and on customer loyalty (ß = -0.260; p = .001; R^2 = 0.231). Therefore, H6a and H6b are supported. There is a statistically insignificant effect of response time on customer satisfaction (ß = 0.039; p = .311; R^2 = 0.263) and customer loyalty (ß = 0.036; p = .317; R^2 = 0.231). Therefore, H7a and H7b are not supported.

The serial mediation effects are shown in Appendix 9C, Table 9C.5. None of the serial mediation paths are statistically significant (< .05). Additionally, the serial mediation effects with moderation are shown in Appendix 9C, Table 9C.6. Corresponding with the serial mediation effects without moderation, none of the serial mediation paths are statistically

significant (< .05). So, H8a and H8b are not supported. Furthermore, the mediation effects between customer confusion, dynamic gaze behavior, and CE-outcomes are shown in Appendix 9C, Table 9C.7. None of the mediation paths are statistically significant (< .05). Moreover, the mediation effects with moderation are shown in Appendix 9C, Table 9C.8. Also here, none of the mediation paths are statistically significant (< .05). Therefore, H9a and H9b are not supported. To continue, the mediation effects between customer confusion, searching for additional information, and CE-outcomes are shown in Appendix 9C, Table 9C.9. There is a statistically significant, negative effect of customer confusion on customer satisfaction, via searching for additional information (ß = -0.172; p = .001; R^2 = 0.263). In addition, there is a statistically significant, negative effect of customer confusion on customer loyalty, via searching for additional information (ß = -0.190; p <.001; R^2 = 0.231). Therefore, H10a and H10b are supported. On the contrary, there is a statistically insignificant effect of customer confusion on satisfaction (ß = 0.040; p = .312; R^2 = 0.263) and on customer loyalty (ß = 0.033; p = .318; R^2 = 0.231), via response time (see Table 9C.9, Appendix 9C) Therefore, H11a and H11b are not supported. Based on the results above, the relationship between customer confusion and customer satisfaction is partially mediated by searching for additional information. The relationship between customer confusion and customer loyalty is fully mediated by searching for additional information.

4.3.4 Additional analysis

The additional analysis included the effects on and of the average duration on non-relevant areas and were included in the structural model to compare relevant and non-relevant areas. There was a statistically insignificant effect of average duration non-relevant areas on searching for additional information (ß = 0.050; p = .249; R^2 = 0.170). Additionally, there was a statistically insignificant effect of average duration non-relevant areas on response time (ß = 0.092; p = .121; R^2 = 0.279). Surprisingly, there is a statistically significant, positive effect of customer confusion on average duration non-relevant areas (ß = 0.385; p = .016; R^2 = 0.014). Furthermore, there was a statistically significant, negative moderation effect of perceived time pressure on the relationship between customer confusion and average duration non-relevant areas (ß = -0.611; p = .019; R^2 = 0.014). This moderation effect is visually illustrated in Appendix 9B, Figure 9B.2. When customers are confused and perceive time pressure, the average duration on non-relevant areas is lower compared to when customers are confused and do not perceive time pressure. So, if a customer is confused and perceives time pressure, the fixations on non-relevant areas are shorter compared to when they are confused but do not

perceive time pressure. Finally, none of the serial mediation or mediation paths showed a statistically significant effect (<.05) (see Appendix 9C, Table 9C.10-9C.12). Furthermore, the additional analysis showed a statistically significant, positive effect of customer satisfaction on customer loyalty (*ß* = 0.632; $p < .001$; $R^2 = 0.231$) and a statistically significant, indirect effect of customer confusion on loyalty, via customer satisfaction (*ß* = -0.169; $p = .049$; $R^2 = 0.231$). Finally, the indirect effect of customer confusion on the behavioral dimension, via the sensorial dimension showed no significant results (*ß* = 0.062; $p = .115$; $R^2 = 0.170$ & *ß* = 0.060; $p = .116$; $R^2 = 0.279$).

Control variables

The path coefficients and corresponding p-values are shown in Appendix 9C, Table 9C.13. Age has a statistically significant, negative effect on loyalty (*ß* = -0.203; $p = .001$; $R^2 = 0.231$) and on satisfaction (*ß* = -0.190; $p = .007$; $R^2 = 0.263$). Confusion avoidance has a statistically significant, negative effect on satisfaction (*ß* = -0.103; $p = .042$; $R^2 = 0.263$). Problem-solving preference has a statistically significant, negative effect on loyalty (*ß* = -0.288; $p < .001$; $R^2 = 0.231$) and on satisfaction (*ß* = -0.300; $p < .001$; $R^2 = 0.263$). The other control variables show a statistically insignificant effect on both loyalty and satisfaction. The whole SmartPLS structural model is shown in Appendix 9C, Figure 9C.3.

An additional, qualitative analysis was done regarding the behavioral dimension to gain more in-depth insights on the content of the messages that were sent during confusion. The qualitative analysis is shown in Appendix 9D. The analysis gave insight into several behavioral responses when individuals are confused. Actions such as asking for or seeking more information were common. Responses such as “asking you where I could find the number you were looking for” and “To ask for help” indicate that individuals required more information. Additionally, restarting the purchase, refreshing the page, quitting the purchase process, or not being willing to buy the product were behavioral intentions mentioned. Finally, a prominent result was that the majority of individuals act intuitively when confused. Responses such as “just providing the different kind of numbers” are indicative of this intuitive behavior, since individuals intuitively find the right information.

5. Conclusion and discussion

The key findings of this research provide insights into how confusion is experienced in a chatbot context, regarding sensations and behavior, and how this confusion affects customer experience outcomes such as satisfaction and loyalty. Specifically, it is examined how confusion affects the dynamic gaze behavior of individuals (also with perceived time pressure as moderator), how confusion affects the search for additional information and response time, and how this experience and the confusion itself affect the satisfaction and loyalty of individuals in chatbot context. Therefore, this research answers the following research question: “How is the relationship between customer confusion and customer satisfaction and loyalty mediated by the customer experience and how is this moderated by perceived time pressure?”. First, the relationship between customer confusion and customer satisfaction is partially mediated by the behavioral dimension of the customer experience. Customer confusion also directly affects customer satisfaction. Additionally, the relationship between customer confusion and customer loyalty is fully mediated by the behavioral dimension of the customer experience. So, customer confusion does not have a direct impact on customer loyalty. The sensorial dimension does not mediate the relationship between customer confusion and customer experience outcomes. Furthermore, customer confusion affects the dynamic gaze behavior of an individual regarding more fixations on relevant areas and a longer duration of fixations on non-relevant areas. With perceived time pressure, this effect is the opposite, resulting in shorter duration on relevant areas and on non-relevant areas. In addition, customer confusion results in the search for additional information and a longer response time. Finally, dynamic gaze behavior has an impact on the search for additional information regarding that more fixations on relevant areas indicate that an individual is searching for additional information.

5.1 Discussion

The results of this research provide an answer to the research question of how the customer experience mediates the relationship between customer confusion and customer experience outcomes and how this is moderated by perceived time pressure. Specifically, insight is gained into how the sensorial and behavioral dimensions of the customer experience influence each other and the relationship between customer confusion and satisfaction, and loyalty in a chatbot context. The most important results can be found in Figure 2.

To begin with, the results provide partial support for the effect of dynamic gaze behavior on relevant areas on searching for additional information. More fixations on relevant areas affect searching for additional information, which is in line with the previous studies by Enders et al. (2021) and He et al. (2014). However, contrary to this study by Enders et al. (2021) is the fact that longer fixations on relevant areas do affect searching for additional information. A reason for this result might be that the performed task was not complex enough to require in-depth information processing. Negi and Mitra (2020) support this by stating that complex tasks require more processing time and longer fixations. In addition, Rayner et al., (2007) state that shorter fixations entail that the information was easy to process. Qualitative research revealed that the provided information was not difficult to process; “It can't be difficult because you have 3 opportunities to give for resolving the error”, however, it was not clear which information was required; “For me, it was not really clear which number I had to give you”. Therefore, a possible explanation for this result is that the participants did not need to process the information in-depth. The additional analysis showed that the average duration of fixations on non-relevant areas did not affect searching for additional information. This is in line with the study of Enders et al. (2021), which states that only gaze behavior on the informative parts of an interface affects the search for information. The results provide no support for the effect of dynamic gaze behavior on response time. A possible explanation for this has to do with the aforementioned experimental design. When processing information in-depth, longer response times were hypothesized (Krajbich et al., 2015; Stupple et al., 2013; Gill and Prowse, 2017). However, since no in-depth information processing was required, dynamic gaze behavior has no effect on response times.

Furthermore, the results provided partial support for the effect of customer confusion on dynamic gaze behavior on relevant areas. Customer confusion results in more fixations on relevant areas to resolve this confusion, which is in line with the study of Salminen et al. (2019). However, contrasting the studies of Cahya et al. (2019) and Salminen et al. (2019) confusion does not have an effect on the average duration of fixations on relevant areas. A possible explanation for this is that confusion was solved intuitively. Qualitative analysis revealed that behavior during confusion can be intuitive when overloaded with information “Just throwing all the number I saw, one by one”, “to give you all the numbers as fast as possible”. Intuitive individuals do not require to fixate longer in order to process information (Ares et al., 2014; Motoki et al., 2021). Therefore, the average duration on relevant areas during confusion does not need to be longer compared to non-confusion. The additional analysis showed that the average duration of fixations on non-relevant areas is longer while

experiencing confusion. This result can be explained by the study of Wachowiak et al. (2022) which states that confused customers look into the environment the most. When individuals encounter a confusing situation during a human-robot interaction, the gaze is more oriented toward the environment than toward the robot (Wachowiak et al., 2022). Therefore, the duration of fixations is longer on non-relevant areas, since individuals look into the environment while experiencing confusion. However, when experiencing confusion, individuals also have the urge to resolve this confusion (Drummond, 2004; Kulthau, 1991). In combination with the significant result of more fixations on relevant areas, it could be argued that confused customers first look into the environment for a while, and then find a way to resolve this confusion by shifting their focus to relevant areas. By shifting their focus onto relevant areas, individuals are more likely to resolve the confusion due to the information that is provided on these relevant areas.

Moving on, the results provided support for the effect of customer confusion on searching for additional information, which is in line with the studies of Kuhlthau (1991), Drummond (2004), and Matzler and Waiguny (2005). When someone is confused, the confusion-reduction strategy of searching for additional information is used to resolve this confusion. Qualitative analysis showed that the confusion-reduction strategy of postponing or abandoning the purchase also belongs to the possibilities of coping with confusion, which corresponds to the study of Sharma et al. (2023). Furthermore, when someone is confused, the response time is longer. This is in line with the study of D'Mello et al. (2014), which states that deliberate cognitive processes are required to resolve confusion. As a consequence, response times are longer due to the cognitive effort that is required (Krajbich et al., 2015; Stupple et al., 2013; Gill and Prowse, 2017). Furthermore, the results provided partial support for an interplay between bottom-up factors (confusion) and top-down factors (time pressure) on dynamic gaze behavior on relevant areas. In a confusing situation with time pressure, the average duration of the fixations on relevant areas is shorter compared to when there is no perceived time pressure. This is in line with the strategy of acceleration to cope with confusion and time pressure explained by Pieters and Warlop (1999). This strategy entails that individuals speed up their visual information search and that therefore the duration of fixations on relevant areas is reduced. However, perceived time pressure does not moderate the relationship between customer confusion and total fixations on relevant areas. Therefore, individuals have the same number of fixations on relevant areas when confused, regardless of whether they perceived time pressure or not. This contradicts the study of Pieter and Warlop (1999), which states that individuals cope with time pressure by fixating less. A possible

explanation for this is that individuals try to resolve their confusion by searching for additional information (Drummond, 2004). In order for individuals to acquire information, the eyes need to be fixated (Meißner & Oll, 2019). This stresses the importance of fixations during confusion, regardless of whether time pressure is perceived or not. The results showed that the average duration of fixations on non-relevant areas is shorter when someone is confused and perceives time pressure, compared to when no time pressure is perceived in a confusing situation. This is in line with the study of Pieter and Warlop (1999) which state that visual information search is even more important when time pressure is perceived. So, when someone perceives time pressure, looking at non-relevant areas will not speed up the process of collecting information during confusion. Therefore, the duration of fixations on non-relevant areas will be shorter when experiencing confusion and time pressure. Comparing both significant moderation effects of time pressure, it can be concluded that time pressure causes confused individuals to fixate shorter on relevant and non-relevant areas.

The results provide evidence for the negative effect of customer confusion on customer satisfaction. This corresponds to previous studies of Moon et al. (2016), Walsh and Mitchell (2010), and Edward and Sahadev (2012). The behavioral dimension of the customer experience partially mediates between customer confusion and customer satisfaction. Contrary to what was hypothesized, the results do not provide evidence for a direct effect of customer confusion on customer loyalty. This result also contradicts the studies of Edward and Sahadev (2012), Reichheld et al. (2020), and Mitchell and Papavassiliou (1999). A possible explanation for this is that customer confusion first affects customer satisfaction, and customer satisfaction affects customer loyalty. In a study by Unurlu (2019), also a mediation effect was shown between confusion and loyalty. Additional analysis provided evidence for the mediation effect of customer satisfaction between customer confusion and customer loyalty. This result is also in line with several studies such as Chang et al. (2009), Choi & Kandampully, 2019, Kusumawati and Rahayu (2020), and Hsu and Lin (2023), which state that a customer first needs to be satisfied before becoming loyal. Another possible explanation is that customer loyalty is mainly affected by the customer experience itself, as supported by the studies of Klaus & Maklan, (2013), Kusumawati & Rahayu, (2020), Zaid & Patwayati, (2021), and Mokha & Kumar, (2022). The results provide support for this since the behavioral dimension fully mediates between customer confusion and customer loyalty.

Furthermore, searching for additional information has a negative effect on customer satisfaction and loyalty, which is in line with Kusumawati & Rahayu (2020) and Mohka and Kumar (2022) who state that a customer experience is perceived as favorable when minimal

time and effort is required. The results do not provide support for the effect of response time on customer satisfaction and loyalty. This result can be explained by the studies of Dörk et al. (2011), Deary et al. (2001), and Pachella, (1973) who state that response time can be considered as the process of visual search (searching for additional information), rather than the actual behavior during an experience. This entails that response time is not the behavioral act with which the customer experience ends, but that the search for additional information is the behavioral act that causes an individual to be dissatisfied and disloyal. So, the results provide partial support for the effect of the behavioral dimension on the CE-outcomes.

The results provide no support for the serial mediation of customer confusion on customer satisfaction and loyalty, via the sensorial and behavioral dimensions. This could be devoted toward the fact that the sensorial dimension alone does not fully affect the behavioral dimension. According to Krishna (2012), Rybanska et al. (2014), Krishna & Schwarz (2014), and Chen & Li (2018), sensations need to work together with emotions and thoughts to form a perception. Behavior is then affected by this perception. Another plausible explanation has to do with the indirect effects of the sensorial dimension. There was no mediation effect of customer confusion on the behavioral dimension, via the sensorial dimension. Additionally, the results provide no support for the mediation of the sensorial dimension between customer confusion and customer experience outcomes. This indicates that sensations have no large enough effect on behavior or on the customer experience outcomes. Although vision is the dominant sense while gaining environmental knowledge (Krishna, 2012; Gallego et al., 2022; Pocock, 1981), all five senses still have to work together in order to create a full understanding of the environment someone is in (Lundborg, 2013). Since vision is the only sense researched here, it could possibly be the case that solely investigating dynamic gaze behavior did not capture the full sensorial experience.

As mentioned above, there was no support for the indirect effect of customer confusion on customer satisfaction and loyalty, via the sensorial dimension. This can be explained by the aforementioned studies as well, stating that sensations alone do not capture the whole experience and therefore do not affect the CE-outcomes. Finally, the results provide partial support for the indirect effect of customer confusion on customer satisfaction and loyalty, via the behavioral dimension. Customer confusion affects customer satisfaction and loyalty via searching for additional information, which is in line with the studies of Klaus & Maklan, (2013), Kusumawati & Rahayu (2020), Zaid & Patwayati (2021), and Mokha & Kumar (2022) stating that the customer experience has an impact on satisfaction and loyalty. Since behavior is the final dimension of the customer experience (Krishna, 2012), this

dimension has a direct effect on these outcomes. Customer confusion has no effect on customer satisfaction and loyalty via response time which can be explained by the fact that response time is not the final behavioral action of the customer experience (Dörk et al., 2011; Deary et al., 2001; Pachella, 1973). The behavioral dimension is a strong mediator between customer confusion and customer satisfaction (partial mediation) and loyalty (full mediation).

To conclude, the results of this research provide (partial) support for the hypothesized effects and also provide insight into new, unexpected effects. Therefore, this research provides more in-depth insights into the relationships between customer confusion, the customer experience, and the customer experience outcomes satisfaction and loyalty.

5.2 Theoretical implications

Although confusion has increasingly become a concern due to the exponentially rising amount of information in the marketplace, the concept received insufficient attention in the marketing literature (Chauhan & Sagar, 2021). Especially with the rapid technological growth and the use of self-service technologies, research on confusion is becoming more important. The few studies that report on confusion, utilize oversimplified measures (Shiu, 2017; Edward & Sahadev, 2012; Walsh & Mitchell, 2010). This demonstrates the necessity of studying confusion from a more holistic perspective to investigate how this confusion impacts the experience of an individual. Subsequently, the present research provides rich insights into how confusion is experienced while interacting with a chatbot, by focusing on the sensorial and behavioral dimension of the customer experience, and how this confusion (together with the experience) ultimately impacts satisfaction and loyalty.

First, due to the lack of proper measurements of the customer experience, research on how the customer experience is influenced and on the consequences of the customer experience is also lacking (Lemon & Verhoef, 2016). Therefore, this research adds to the existing literature on the drivers and consequences of the customer experience by focusing on confusion as a driver and the interaction between confusion and perceived time pressure as a driver. In addition, this research investigates two customer experience outcomes, satisfaction, and loyalty. The key findings demonstrate that confusion affects both the sensorial and behavioral dimensions of the customer experience. Furthermore, findings show that confusion impacts satisfaction directly, and indirectly via the behavioral dimension. In contrast, confusion impacts loyalty only via the customer experience, and not directly. To conclude, this research contributes to the existing literature by focusing on confusion as a driver of the

customer experience and showing how it impacts the experience of an individual, also under perceived time pressure. In contrast to other studies (e.g., Shiu, 2017; Edward & Sahadev, 2012; Walsh & Mitchell, 2010), which use oversimplified measures of confusion, this research examines how confusion affects the experience of an individual by empirically testing it, rather than measuring confusion on a scale. Therefore, this research studies confusion from a more holistic perspective and thereby adds value to the measurement of confusion. In extension, in contrast to other studies (e.g., Edward & Sahadev, 2012; Moon et al., 2016), this research shows how confusion has an impact on satisfaction and loyalty, through the customer experience. Therefore, this research adds value by gaining insight into how satisfaction and loyalty are both affected by confusion and the customer experience.

Second, due to the multidimensionality of the customer experience, more research is needed on how to empirically measure the complex concept and its dimensions (Jain et al., 2017; Hwang & Seo, 2016; Lemon & Verhoef, 2016). Therefore, this research contributes to the existing literature on how the customer experience and its dimensions can be empirically measured by showing how confusion impacts the sensorial and behavioral dimensions. By the use of an eye-tracker, the visual processing part of the sensorial dimension can be accurately measured. Key findings show that confusion results in more dynamic gaze behavior in terms of fixations on relevant areas and longer fixations on non-relevant areas. Derived from that, individuals first fixate on the environment when confused, to then fixate on informative parts to resolve the confusion. Additionally, by examining the real-time behavior of an individual during a chatbot experiment, and analyzing text on behavioral intentions, insights can be gained on how an individual behaves when confused. Key findings show that confusion results in searching for additional information, and a longer response time. To conclude, this research adds to the existing literature on how to empirically measure the customer experience and its dimensions by utilizing an eye-tracker to measure sensations and by tracking real-time behavior to measure the behavioral aspect of the customer experience. Therefore, this research shows that the dimensions of the customer experience can be measured simultaneously by means of a combination of an eye-tracker and a chatbot, which aligns with the required multidimensional view proposed by Lemon and Verhoef (2016). Whereas studies such as Brun et al. (2017), and Bagdare and Jain (2013) measure the customer experience dimensions in isolation, this research shows that – by utilizing eye-tracking and chatbots – the dimensions can be measured while maintaining their interrelations.

Third and last, research on the sensorial dimension is scarce, despite the importance of this dimension (Mahr et al., 2019; Bustamante & Rubio, 2017; Keiningham et al., 2017). Additionally, Gerea et al. (2021) address the importance of behavioral data in order to predict future behavior. Furthermore, Lemon and Verhoef (2016) address the importance of investigating the customer experience dimensions simultaneously. Therefore, this research contributes to the existing literature by focusing explicitly on these two dimensions which need to be examined the most. Furthermore, the interrelation between these dimensions is investigated, to see how the visual processing process impacts someone's behavior. Empirical data on the senses and behavioral data are incorporated to reach a conclusion about these dimensions. Key findings show that the total fixations on relevant areas affect the search for additional information. However, this finding is too fragile to state that the sensorial dimension directly impacts the behavioral dimension. Despite this, it does show that these dimensions of the customer experience interrelate in some way. To conclude, this research adds to the existing literature by incorporating empirical data on the senses and behavioral data to draw a conclusion about how these dimensions interrelate. By utilizing an eye-tracker, in-depth insights into an individual's gaze behavior are provided, addressing the importance and need for empirical data on the senses (Mahr et al., 2019). Furthermore, by tracking real-time behavior and analyzing behavioral intentions by means of qualitative research, valuable insights into the behavioral dimension are gained, such as how this affects satisfaction and loyalty, which aligns with the proposed view of Gerea et al. (2021). The combination of behavioral and attitudinal data contributes to the ability to predict future behavior.

5.3 Practical and managerial implications

Ultimately, chatbots are designed to provide customer service activities with the intention of increasing customer satisfaction (Nicolescu & Tudorache, 2022). However, interacting with a chatbot can be confusing at times for customers (Lu & Gursoy, 2014). This confusion is then experienced as negative and this will impact satisfaction and loyalty (Ma et al., 2023; Nicolescu & Tudorache, 2022). Eventually, the disadvantages of confusion will undermine the potential 67% increase in sales that chatbots can deliver (Beckmann, 2022). To dive deeper into how this confusion is experienced in a chatbot context and how firms can then respond to this confusion, this research examines the effects of confusion on the sensorial and behavioral dimensions of the customer experience. Additionally, it is examined how confusion impacts satisfaction and loyalty, via the customer experience in order for firms to

pick up signals that are indicative of confusion. Therefore, this research contributes to the practical problems encountered in firms by providing an in-depth examination of sensorial and behavioral responses that are indicative of confusion.

First, when a customer is confused, this results in more fixations on relevant areas and a longer duration of fixations on non-relevant areas. Therefore, the viewing pattern is quite spread out when an individual is confused. This result gives managers the ability to recognize confusion with the help of a heatmap since this is a useful tool to examine whether gaze behavior is spread out (Djamasbi, 2014). Since the eye-tracking technology has not yet been sufficiently developed, heatmaps are only available for mouse movements and the number of clicks on a webpage. However, previous research by Papoutsaki et al. (2018), Liebling and Dumais (2014), Chen et al. (2001), and Huang et al. (2012) state that mouse movements or mouse clicks have a strong relationship with gaze behavior during human-computer interaction. Specifically, the mouse had been characterized as an additional eye-tracker. When the gaze position is fixated on a part of the screen, it is likely that the cursor is also located in that part of the screen. Additionally, before clicking, first the eyes fixate on the location of the click (Papoutsaki et al., 2018). To conclude, heatmaps are able to show if a customer gets confused when browsing a webpage. As a result, managers can recognize confusion through heatmaps and check whether or not the webpage is successfully designed (Djamasbi, 2014). When the heatmaps show a spread-out pattern, then this is indicative of a confused user and firms can respond to this by adjusting the webpage.

Second, key findings show that the satisfaction of a customer is directly impacted by confusion, and that loyalty is impacted via the unfavorable customer experience caused by confusion. Recalling the disadvantages of confusion on firm outcomes, it is crucial for firms to prevent confusion from happening and this is also one of the major aims of firms (Shukla et al., 2010). Especially in the financial service industry or in the tourism industry, where similar, ambiguous, and an overload of offerings are more the norm than the exception, prevention of confusion is a potential source of competitive advantage (Shukla et al., 2010; Park & Jang, 2013). In this research, confusion was manipulated by providing ambiguous, and an overload of information. Therefore, preventing confusion starts with building chatbots and/or webpages that provide clear, straightforward, and simple information (Ranjbarian et al. 2016). Additionally, when building chatbots and/or webpages, managers should avoid ambiguous or misleading words, phrases, or descriptions that could cause customers to misinterpret information (Lu et al., 2016). Furthermore, managers must be mindful of the amount of information that is provided (Ranjbarian et al., 2016; Lu et al., 2016). This is

especially important in situations with time pressure since individuals then need to find the right information in a shorter amount of time looking at the online environment.

The discussion above reveals that confusion can be caused by many aspects of information, which is in line with the research of Shukla et al. (2010). Therefore, managers should not focus on confusion as a unidimensional construct but rather should treat confusion from multiple perspectives. As Shukla et al. (2010) mention, confusion about the features of products can be prevented by segmenting the products on high-, medium-, and low-quality features. This segmenting can be done by conducting tests, and by testing products in the marketplace with customers. However, unclear or low-quality information might still lead to confusion. Additionally, Matzler et al. (2011) argue that the characteristics of individuals determine whether or not an individual can handle unclear, ambiguous, and an overload of information. Specifically, knowledgeable customers are less likely to become confused by this since they have clear wants and needs and can filter out the information that they need. Therefore, to prevent confusion from happening, a customer segmentation strategy is an appropriate solution. For example, tailoring webpages to different groups of customers with different levels of experience or product knowledge could help prevent confusion. A small introductory assessment of customer's knowledge or experience by asking simple questions before entering a webpage gives firms the ability to forward customers to a tailored webpage. For example, a webpage with more options should be offered to more knowledgeable customers in order for them to choose the best option based on their wants and needs.

In conclusion, due to this research, managers can recognize confusion in an early stage of the experience by utilizing heat maps and analyzing whether or not the viewing pattern is spread out. Responding accordingly to this confusion in an early stage is crucial because the disadvantages of confusion might be avoided then. Recognizing confusion is important, however, preventing it is even better. This research shows that this starts with providing clear, straightforward, simple, and unambiguous information in the right quantity. A step further, however, is for managers to tailor webpages to different groups of customers, based on individual characteristics (e.g., product knowledge), which is derived from a small assessment before entering a webpage. The more managers focus on confusion from multiple perspectives, the better they will in optimizing the organizational resources to cope with or prevent confusion.

5.4 Limitations and further research

The present study provides in-depth insights into how individuals experience confusion regarding their senses and their behavior, and how this impacts their satisfaction and loyalty. However, this research also has limitations. First, this research solely focused on confusion in a chatbot context, while there are many more self-service technologies that might cause confusion (Mendat & Mayhorn, 2007). Therefore, further research should investigate confusion caused by other self-service technologies to get a more complete picture. Second, this research focused on two dimensions of the customer experience, while the customer experience officially encompasses five dimensions. Due to this, the customer experience is not studied in its entirety. Following this, the only human sense that was examined is vision. Chances are, therefore, that the sensorial dimension is not examined in its entirety. So, further research on the customer experience dimensions is needed to study this complex concept in its entirety and to see how the dimensions work together. Especially the sensorial dimension needs further research. This dimension is the most valuable dimension of the customer experience (Gentile et al., 2007; Brakus et al., 2009), however, is has not been studied extensively yet (Mahr et al., 2019; Gerea et al., 2021). Further research should aim to focus on all the senses since these all work together to form a perception of the surroundings.

Third, convenience sampling is the sampling technique that was utilized. Therefore, sampling bias might be present. For generalizability, further research should aim to randomize the sample. Fourth, the eye-tracker broke down during the preparation of the experiment. Due to this, the right eye-camera could not be used, and thus are the eye-tracking data solely based on the left eye-camera. As a result, some angles of the screen could not be reached by the eye-tracker. This could have had a large impact on the results and therefore should be taken into account when interpreting the results. Further research should include a complete, working eye-tracker to rule out the possibility that the results are affected by technological defects. Fifth, the majority of the constructs used in the research were single-item constructs. In general, single-item constructs are not recommended because then the reliability of the constructs is not known (Petrescu, 2013). These single-item constructs are indicative of poor measurements of the constructs. Therefore, further research should provide better measurements for the constructs to avoid having single-item constructs. Finally, time limitations caused that dynamic gaze behavior could not be studied more in-depth, as well as the behavioral dimension. So, further research should focus on studying the sensorial and behavioral dimensions more in-depth including more eye-movement metrics, facial expressions, mouse movements, and/or body movements.

References

Abd-Alrazaq, A., Alajlani, M., Ali, N., Denecke, K., Bewick, B. M., & Househ, M. (2021). Perceptions and Opinions of Patients About Mental Health Chatbots: Scoping Review. *Journal of Medical Internet Research*, *23*(1), e17828. https://doi.org/10.2196/17828

Achjari, D. (2015c). Partial Least Squares: Another Method of Structural Equation Modeling Analysis. *Journal of Indonesian Economy and Business*, *19*(3). https://doi.org/10.22146/jieb.6599

Adamopoulou, E., & Moussiades, L. (2020). Chatbots: History, technology, and applications. *Machine learning with applications*, *2*, 100006. https://doi.org/10.1016/j.mlwa.2020.100006

Agapito, D., Mendes, J., & Valle, P. O. D. (2013). Exploring the conceptualization of the sensory dimension of tourist experiences. *Journal of Destination Marketing and Management*, *2*(2), 62–73. https://doi.org/10.1016/j.jdmm.2013.03.001

Akbari, M., Kazemi, R., & Haddadi, M. (2016). Relationship marketing and word-of-mouth communications: Examining the mediating role of customer loyalty. *Marketing and branding research*, *3*(1), 63–74. https://doi.org/10.33844/mbr.2016.60193

Åkesson, M., Edvardsson, B., & Tronvoll, B. (2014). Customer experience from a self-service system perspective. *Journal of Service Management*, *25*(5), 677–698. https://doi.org/10.1108/josm-01-2013-0016

Anderson, C. J. (2010). Central Limit Theorem. *Corsini Encyclopedia of Psychology*, 1–2. https://doi.org/10.1002/9780470479216.corpsy0160

Anninou, I., & Foxall, G. R. (2019). The reinforcing and aversive consequences of customer experience. The role of consumer confusion. *Journal of Retailing and Consumer Services*, *51*, 139–151. https://doi.org/10.1016/j.jretconser.2019.05.029

Ares, G., Mawad, F., Giménez, A., & Maiche, A. (2014). Influence of rational and intuitive thinking styles on food choice: Preliminary evidence from an eye-tracking study with yogurt labels. *Food Quality and Preference*, *31*, 28–37. https://doi.org/10.1016/j.foodqual.2013.07.005

Asnawi, A. L., Gravell, A. M., & Wills, G. (2012). *Factor Analysis: Investigating Important Aspects for Agile Adoption in Malaysia*. https://doi.org/10.1109/agileindia.2012.13

Ballard, D. H., Hayhoe, M., & Pelz, J. B. (1995). Memory Representations in Natural Tasks. *Journal of Cognitive Neuroscience*, *7*(1), 66–80. https://doi.org/10.1162/jocn.1995.7.1.66

Beckmann, J. (2023, 14 februari). *23 Chatbot Statistics for 2023*. Simplr. https://www.simplr.ai/blog/22-chatbot-statistics-for-2022

Brakus, J. J., Schmitt, B. H., & Zarantonello, L. (2009). Brand Experience: What Is It? How Is It Measured? Does It Affect Loyalty? *Journal of Marketing*, *73*(3), 52–68. https://doi.org/10.1509/jmkg.73.3.52

Brandtzaeg, P., & Følstad, A. (2017). Why People Use Chatbots. In *Lecture Notes in Computer Science* (pp. 377–392). Springer Science+Business Media. https://doi.org/10.1007/978-3-319-70284-1_30

Brun, I., Rajaobelina, L., Ricard, L., & Berthiaume, B. (2017). Impact of customer experience on loyalty: a multichannel examination. *Service Industries Journal*, *37*(5–6), 317–340. https://doi.org/10.1080/02642069.2017.1322959

Business Wire. (2022, 16 September). *The Worldwide Chatbot Industry is Expected to Reach $22.9 Billion by 2030 - ResearchAndMarkets.com.* https://www.businesswire.com/news/home/20220916005448/en/The-Worldwide-Chatbot-Industry-is-Expected-to-Reach-22.9-Billion-by-2030---ResearchAndMarkets.com

Bustamante, J. M., & Rubio, N. (2017). Measuring customer experience in physical retail environments. *Journal of Service Management*, *28*(5), 884–913. https://doi.org/10.1108/josm-06-2016-0142

Chan, C. Y. T., & Petrikat, D. (2022). Self-Service Technology: Benefits and Challenges. *Journal of computer science and technology studies*, *4*(2), 118–127. https://doi.org/10.32996/jcsts.2022.4.2.14

Chang, H. Y., Wang, Y., & Yang, W. (2009). The impact of e-service quality, customer satisfaction and loyalty on e-marketing: Moderating effect of perceived value. *Total quality management & business excellence*, *20*(4), 423–443. https://doi.org/10.1080/14783360902781923

Chauhan, V., & Sagar, M. (2021). Consumer confusion: a systematic review and research directions. *Journal of Consumer Marketing*, *38*(4), 445–456. https://doi.org/10.1108/jcm-03-2020-3705

Chen, H., & Lin, Y. (2018). A study of the relationships among sensory experience, emotion, and buying behavior in coffeehouse chains. *Service Business*, *12*(3), 551–573. https://doi.org/10.1007/s11628-017-0354-5

Chen, M., Anderson, J. R., & Sohn, M. (2001). *What can a mouse cursor tell us more?* https://doi.org/10.1145/634067.634234

Chen, N., Mohanty, S., Jiao, J., & Fan, X. (2021). To err is human: Tolerate humans instead of machines in service failure. *Journal of Retailing and Consumer Services*, *59*, 102363. https://doi.org/10.1016/j.jretconser.2020.102363

Choi, H., & Kandampully, J. (2019). The effect of atmosphere on customer engagement in upscale hotels: An application of S-O-R paradigm. *International Journal of Hospitality Management*, *77*, 40–50. https://doi.org/10.1016/j.ijhm.2018.06.012

Chung, M., Ko, E., Joung, H., & Kim, S. H. (2020). Chatbot e-service and customer satisfaction regarding luxury brands. *Journal of Business Research*, *117*, 587–595. https://doi.org/10.1016/j.jbusres.2018.10.004

Dayan, P., Kakade, S. M., & Montague, P. R. (2000). Learning and selective attention. *Nature Neuroscience*, *3*(S11), 1218–1223. https://doi.org/10.1038/81504

Dash, G., & Paul, J. (2021). CB-SEM vs PLS-SEM methods for research in social sciences and technology forecasting. *Technological Forecasting and Social Change*, *173*, 121092. https://doi.org/10.1016/j.techfore.2021.121092

De Dreu, C. K. W. (2003). Time pressure and closing of the mind in negotiation. *Organizational Behavior and Human Decision Processes*, *91*(2), 280–295. https://doi.org/10.1016/s0749-5978(03)00022-0

Djamasbi, S. (2014). Eye Tracking and Web Experience. *AIS transactions on human-computer interaction*, *6*(2), 37–54. https://doi.org/10.17705/1thci.00060

D'Mello, S. K., Lehman, B., Pekrun, R., & Graesser, A. C. (2014). Confusion can be beneficial for learning. *Learning and Instruction*, *29*, 153–170. https://doi.org/10.1016/j.learninstruc.2012.05.003

Dörk, M., Carpendale, S., & Williamson, C. (2011, May). The information flaneur: A fresh look at information seeking. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1215-1224). https://doi.org/10.1145/1978942.1979124

Drummond, G. (2004). Consumer confusion: reduction strategies in higher education. *International Journal of Educational Management*, *18*(5), 317–323. https://doi.org/10.1108/09513540410543466

Edward, M., & Sahadev, S. (2012). Modeling the consequences of customer confusion in a service marketing context: An empirical study. *Journal of Services Research*, *12*(2).

Emerson, R. W. (2015). Convenience Sampling, Random Sampling, and Snowball Sampling: How Does Sampling Affect the Validity of Research? *Journal of Visual Impairment & Blindness*, *109*(2), 164–168. https://doi.org/10.1177/0145482x1510900215

Enders, L. R., Smith, R., Gordon, S. B., Ries, A. J., & Touryan, J. (2021). Gaze Behavior During Navigation and Visual Search of an Open-World Virtual Environment. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.681042

Eskiler, E., & Safak, F. (2022). Effect of Customer Experience Quality on Loyalty in Fitness Services. *Physical Culture and Sport. Studies and Research*, *94*(1), 21–34. https://doi.org/10.2478/pcssr-2022-0003

Field, A. P. (2017). *Discovering Statistics Using Ibm Spss Statistics*. https://dl.acm.org/citation.cfm?id=2502692

Fitzgerald, M. P., Donovan, K. R., Kees, J., & Kozup, J. (2019). How confusion impacts product labeling perceptions. *Journal of Consumer Marketing*, *36*(2), 306–316. https://doi.org/10.1108/jcm-08-2017-2307

Fotheringham, D., & Wiles, M. V. (2022). The effect of implementing chatbot customer service on stock returns: an event study analysis. *Journal of the Academy of Marketing Science*. https://doi.org/10.1007/s11747-022-00841-2

Gallego, G., Delbruck, T., Orchard, G., Bartolozzi, C., Taba, B., Censi, A., Leutenegger, S., Davison, A. J., Conradt, J., Kumar, V., & Scaramuzza, D. (2022). Event-Based Vision: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *44*(1), 154–180. https://doi.org/10.1109/tpami.2020.3008413

Garaus, M., & Wagner, U. (2016). Retail shopper confusion: Conceptualization, scale development, and consequences. *Journal of Business Research*, *69*(9), 3459–3467. https://doi.org/10.1016/j.jbusres.2016.01.040

Gentile, C., Spiller, N., & Noci, G. (2007). How to Sustain the Customer Experience: *European Management Journal*, *25*(5), 395–410. https://doi.org/10.1016/j.emj.2007.08.005

Gerea, C., Gonzalez-Lopez, F., & Herskovic, V. (2021). Omnichannel Customer Experience and Management: An Integrative Review and Research Agenda. *Sustainability*, *13*(5), 2824. https://doi.org/10.3390/su13052824

Gill, D., & Prowse, V. L. (2017). Using Response Times to Measure Strategic Complexity and the Value of Thinking in Games. *Social Science Research Network*. https://doi.org/10.2139/ssrn.2902411

Goeschl, T., & Lohse, J. (2018). Cooperation in public good games. Calculated or confused? *European Economic Review*, *107*, 185–203. https://doi.org/10.1016/j.euroecorev.2018.05.007

Granka, L., Feusner, M. K., & Lorigo, L. (2008). Eye Monitoring in Online Search. *Springer Berlin Heidelberg eBooks*, 347–372. https://doi.org/10.1007/978-3-540-75412-1_16

Gromski, P. S., Xu, Y., Kotze, H. L., Correa, E., Ellis, D., Armitage, E. G., Turner, M. S., & Goodacre, R. (2014). Influence of Missing Values Substitutes on Multivariate Analysis of Metabolomics Data. *Metabolites*, *4*(2), 433–452. https://doi.org/10.3390/metabo4020433

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

He, J., Siu, C. Y. S., Chaparro, B. S., & Strohl, J. (2014). Mobile. In *Elsevier eBooks* (pp. 255–290). https://doi.org/10.1016/b978-0-12-408138-3.00010-8

Heinzl, H., & Mittlböck, M. (2002). Adjusted R2 Measures for the Inverse Gaussian Regression Model. *Computational Statistics*, *17*(4), 525–544. https://doi.org/10.1007/s001800200125

Hoque, M. Z., Bashar, M. K., & Akhter, F. (2021). Ambiguity tolerance and confusion avoidance in the intent to purchase farmed fish. *Journal of Applied Aquaculture*, *34*(3), 564–597. https://doi.org/10.1080/10454438.2021.1872462

Hsu, C. L., & Lin, J. C. (2023). Understanding the user satisfaction and loyalty of customer service chatbots. *Journal of Retailing and Consumer Services*, *71*, 103211. https://doi.org/10.1016/j.jretconser.2022.103211

Huang, J., White, R. W., & Buscher, G. (2012). *User see, user point*. https://doi.org/10.1145/2207676.2208591

Hui, Z., Mualla, Y., Tchappi, I. H., Galland, S., Bellemans, T., & Yasar, A. (2020). Decision-Making under Time Pressure when Rescheduling Daily Activities. *Procedia Computer Science*, *170*, 281–288. https://doi.org/10.1016/j.procs.2020.03.041

Hwang, J., & Seo, S. (2016). A critical review of research on customer experience management. *International Journal of Contemporary Hospitality Management*, *28*(10), 2218–2246. https://doi.org/10.1108/ijchm-04-2015-0192

IBM Corp. (2021). IBM SPSS Statistics for Windows, Version 28.0. Armonk, NY: IBM Corp

Itier, R. J., & Batty, M. (2009). Neural bases of eye and gaze processing: The core of social cognition. *Neuroscience & Biobehavioral Reviews*, *33*(6), 843–863. https://doi.org/10.1016/j.neubiorev.2009.02.004

Iyer, E. S. (1989). Unplanned purchasing: Knowledge of shopping environment and time pressure. *Journal of Retailing*. https://psycnet.apa.org/record/1989-41539-001

Jacobs, E. & Korzilius H. (2022). *'SPSS-uitwerking Regressie deel 1'.* https://brightspace.ru.nl/d2l/le/content/261736/viewContent/1627667/View

Jain, R., Aagja, J. P., & Bagdare, S. (2017). Customer experience – a review and research agenda. *Journal of service theory and practice*, *27*(3), 642–662. https://doi.org/10.1108/jstp-03-2015-0064

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87*(4), 329–354. https://doi.org/10.1037/0033-295x.87.4.329

Kahneman, D. (2011). *Thinking, fast and slow.* Farrar, Straus and Giroux.

Kannengiesser, U., & Gero, J. S. (2019). Design thinking, fast and slow: A framework for Kahneman's dual-system theory in design. *Design science*, *5*. https://doi.org/10.1017/dsj.2019.9

Keiningham, T. L., Ball, J., Benoit, S., Bruce, H., Buoye, A., Dzenkovska, J., Nasr, L., Ou, Y., & Zaki, M. J. (2017). The interplay of customer experience and commitment. *Journal of Services Marketing*, *31*(2), 148–160. https://doi.org/10.1108/jsm-09-2016-0337

Klaus, P., & Maklan, S. (2013). Towards a Better Measure of Customer Experience. *International Journal of Market Research*, *55*(2), 227–246. https://doi.org/10.2501/ijmr-2013-021

KNAW; NFU; NOW; TO2-Federation; Netherlands Association of Universities of Applied Sciences; VSNU (2018): Netherlands Code of Conduct for Research Integrity.

Kokkinou, A., & Cranage, D. A. (2013). Using self-service technology to reduce customer waiting times. *International Journal of Hospitality Management*, *33*, 435–445. https://doi.org/10.1016/j.ijhm.2012.11.003

Kolb, H. (2003). How the Retina Works: Much of the construction of an image takes place in the retina itself through the use of specialized neural circuits. *American Scientist*, *91*(1), 28–35. http://www.jstor.org/stable/27858152

Krajbich, I., Bartling, B., Hare, T. A., & Fehr, E. (2015). Rethinking fast and slow based on a critique of reaction-time reverse inference. *Nature Communications*, *6*(1). https://doi.org/10.1038/ncomms8455

Krishna, A. (2012). An integrative review of sensory marketing: Engaging the senses to affect perception, judgment and behavior. *Journal of Consumer Psychology*, *22*(3), 332–351. https://doi.org/10.1016/j.jcps.2011.08.003

Krishna, A., & Schwarz, N. (2014). Sensory marketing, embodiment, and grounded cognition: A review and introduction. *Journal of Consumer Psychology*, *24*(2), 159–168. https://doi.org/10.1016/j.jcps.2013.12.006

Kuhlthau, C. C. (1991). Inside the search process: Information seeking from the user's perspective. *Journal of the American Society for Information Science*, *42*(5), 361–371. https://doi.org/10.1002/(sici)1097-4571(199106)42:5

Kusumawati, A., & Sri Rahayu, K. (2020). The effect of experience quality on customer perceived value and customer satisfaction and its impact on customer loyalty. *Human Systems Management*, *39*(2), 219–232. https://doi.org/10.3233/hsm-190564

Lallé, S., Conati, C., & Carenini, G. (2016). Predicting confusion in information visualization from eye tracking and interaction data. *International Joint Conference on Artificial Intelligence*, 2529–2535. http://dblp.uni-trier.de/db/conf/ijcai/ijcai2016.html#LalleCC16

Lee, J. Y., & Huber, J. C. (2021). Evaluation of Multiple Imputation with Large Proportions of Missing Data: How Much Is Too Much? *Iranian journal of public health*. https://doi.org/10.18502/ijph.v50i7.6626

Lee, H., & Yang, K. (2013). Interpersonal service quality, self-service technology (SST) service quality, and retail patronage. *Journal of Retailing and Consumer Services*, *20*(1), 51–57. https://doi.org/10.1016/j.jretconser.2012.10.005

Leek, S., & Kun, D. (2006). Consumer confusion in the Chinese personal computer market. *Journal of Product & Brand Management*, *15*(3), 184–193. https://doi.org/10.1108/10610420610668621

Lemon, K. N., & Verhoef, P. C. (2016). Understanding Customer Experience Throughout the Customer Journey. *Journal of Marketing*, *80*(6), 69–96. https://doi.org/10.1509/jm.15.0420

Liebling, D. J., & Dumais, S. T. (2014). *Gaze and mouse coordination in everyday work*. https://doi.org/10.1145/2638728.2641692

Lu, A. C. C., & Gursoy, D. (2015). A conceptual model of consumers' online tourism confusion. *International Journal of Contemporary Hospitality Management*, *27*(6), 1320–1342. https://doi.org/10.1108/ijchm-04-2014-0171

Lu, A. C. C., Gursoy, D., & Lu, C. Y. (2016b). Antecedents and outcomes of consumers' confusion in the online tourism domain. *Annals of Tourism Research*, *57*, 76–93. https://doi.org/10.1016/j.annals.2015.11.020

Lundborg, G. (2013). The Interaction of the Senses. In *Springer eBooks* (pp. 103–115). https://doi.org/10.1007/978-1-4471-5334-4_10

Ma, H., Kwon, J., & Ahn, J. (2023). The impact of regulations and motivations on behavioural intentions of customers with self-service technology difficulties. *Current Issues in Tourism*, 1–12. https://doi.org/10.1080/13683500.2023.2173056

Mahr, D., Stead, S., & Odekerken-Schröder, G. (2019). Making sense of customer service experiences: a text mining review. *Journal of Services Marketing*, *33*(1), 88–103. https://doi.org/10.1108/jsm-10-2018-0295

Marfo, J. S., Owusu-Bio, M. K., & Asamoah, P. (2022). The Effect of Self-Service Technologies on Customer Experiences in Banking. *International Journal of Online Marketing*, *12*(1), 1–18. https://doi.org/10.4018/ijom.299404

Maruping, L. M., Venkatesh, V., Thatcher, S. M. B., & Patel, P. C. (2015). Folding Under Pressure or Rising to the Occasion? Perceived Time Pressure and the Moderating Role of Team Temporal Leadership. *Academy of Management Journal*, *58*(5), 1313–1333. https://doi.org/10.5465/amj.2012.0468

Matzler, Kurt & Waiguny, Martin. (2005). Consequences of Customer Confusion in Online Hotel Booking. Information and Communication Technologies in Tourism. 10. 306-317. 10.1007/3-211-27283-6_28.

Matzler, K., Stieger, D., & Füller, J. (2011). Consumer Confusion in Internet-Based Mass Customization: Testing a Network of Antecedents and Consequences. *Journal of Consumer Policy*, *34*(2), 231–247. https://doi.org/10.1007/s10603-011-9157-6

McLain, D. L. (2009). Evidence of the properties of an ambiguity tolerance measure: The multiple stimulus types ambiguity tolerance scale–II (MSTAT–II). *Psychological reports*, *105*(3), 975-988.

Meißner, M., & Oll, J. (2019). The Promise of Eye-Tracking Methodology in Organizational Research: A Taxonomy, Review, and Future Avenues. *Organizational Research Methods*, *22*(2), 590–617. https://doi.org/10.1177/1094428117744882

Mendat, C. C., & Mayhorn, C. B. (2007). An Evaluation of Self-Checkout Systems. *Proceedings of the Human Factors and Ergonomics Society . . . Annual Meeting*. https://doi.org/10.1177/154193120705101703

Meuter, M. L., Bitner, M. J., Ostrom, A. L., & Brown, S. J. (2005). Choosing among Alternative Service Delivery Modes: An Investigation of Customer Trial of Self-Service Technologies. *Journal of Marketing*, *69*(2), 61–83. https://doi.org/10.1509/jmkg.69.2.61.60759

Meuter, M. L., Ostrom, A. L., Roundtree, R. I., & Bitner, M. J. (2000). Self-Service Technologies: Understanding Customer Satisfaction with Technology-Based Service Encounters. *Journal of Marketing*, *64*(3), 50–64. https://doi.org/10.1509/jmkg.64.3.50.18024

Meyer, C., & Schwager, A. (2007). Understanding customer experience. *Harvard Business Review*, *85*(2), 116–126, 157.

Mitchell, V., & Papavassiliou, V. G. (1997). Exploring consumer confusion in the watch market. *Marketing Intelligence & Planning*, *15*(4), 164–172. https://doi.org/10.1108/02634509710185270

Mitchell, V., & Papavassiliou, V. (1999). Marketing causes and implications of consumer confusion. *Journal of Product & Brand Management*, *8*(4), 319–342. https://doi.org/10.1108/10610429910284300

Mitchell, V., Walsh, G., & Yamin, M. (2005). Towards a Conceptual Model of Consumer Confusion. *Advances in Consumer Research*, *32*, 143–150. http://www.acrwebsite.org/volumes/v32/acr_vol32_68.pdf

Mokha, A. K., & Kumar, P. (2021). Examining the Interconnections Between E-CRM, Customer Experience, Customer Satisfaction and Customer Loyalty. *Journal of Electronic Commerce in Organizations*, *20*(1), 1–21. https://doi.org/10.4018/jeco.292474

Moldes, O. (2022). Beyond experiential spending: Consumers report higher well-being from purchases that satisfy intrinsic goals. *British Journal of Social Psychology*. https://doi.org/10.1111/bjso.12602

Moon, S. J., Costello, J. P., & Koo, D. M. (2016). The impact of consumer confusion from eco-labels on negative WOM, distrust, and dissatisfaction. *International Journal of Advertising*, *36*(2), 246–271. https://doi.org/10.1080/02650487.2016.1158223

Motoki, K., Saito, T., & Onuma, T. (2021). Eye-tracking research on sensory and consumer science: A review, pitfalls and future directions. *Food Research International*, *145*, 110389. https://doi.org/10.1016/j.foodres.2021.110389

Murdoch, J. N., Murphy, G. C., Alder, R., Gilbert, J., Fierlbeck, K., & Steenbeek, A. (2018). Readiness to Collaborate Scale: How ready are obstetrical practitioners to participate in an interprofessional healthcare team? *Journal of Interprofessional Education and Practice*. https://doi.org/10.1016/j.xjep.2018.09.004

Negi, S., & Mitra, R. (2020). Fixation duration and the learning process: an eye tracking study with subtitled videos. *Journal of Eye Movement Research*, *13*(6). https://doi.org/10.16910/jemr.13.6.1

Nicolescu, L., & Tudorache, M. (2022). Human-Computer Interaction in Customer Service: The Experience with AI Chatbots—A Systematic Literature Review. *Electronics*, *11*(10), 1579. https://doi.org/10.3390/electronics11101579

Noton, D., & Stark, L. (1971). Eye Movements and Visual Perception. *Scientific American*, *224*(6), 34–43. http://www.jstor.org/stable/24922750

Oh, H., Jeong, M., & Baloglu, S. (2013). Tourists' adoption of self-service technologies at resort hotels. *Journal of Business Research*, *66*(6), 692–699. https://doi.org/10.1016/j.jbusres.2011.09.005

Ordóñez, L. D., Benson, L., & Pittarello, A. (2015). Time-pressure Perception and Decision Making. In *John Wiley & Sons, Ltd eBooks* (pp. 517–542). https://doi.org/10.1002/9781118468333.ch18

Orquin, J. L., & Loose, S. M. (2013). Attention and choice: A review on eye movements in decision making. *Acta Psychologica*, *144*(1), 190–206. https://doi.org/10.1016/j.actpsy.2013.06.003

Orquin, J. L., Bagger, M. P., & Loose, S. M. (2013). Learning affects top down and bottom-up modulation of eye movements in decision making. *Judgment and Decision Making*, *8*(6), 700–716. https://doi.org/10.1017/s1930297500004733

Otterbring, T., Wästlund, E., & Gustafsson, A. (2016). Eye-tracking customers' visual attention in the wild: Dynamic gaze behavior moderates the effect of store familiarity on navigational fluency. *Journal of Retailing and Consumer Services*, *28*, 165–170. https://doi.org/10.1016/j.jretconser.2015.09.004

Pachella, Robert. (1973). The Interpretation of Reaction Time in Information Processing Research. Human information processing. 94.

Pachman, M., Arguel, A., Lockyer, L., Kennedy, G., & Lodge, J. M. (2016). Eye tracking and early detection of confusion in digital learning environments: Proof of concept. *Australasian Journal of Educational Technology*, *32*(6). https://doi.org/10.14742/ajet.3060

Papoutsaki, A., Gokaslan, A., Tompkin, J., He, Y., & Huang, J. (2018). The eye of the typer. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*. https://doi.org/10.1145/3204493.3204552

Park, J., & Jang, S. (2013). Confused by too many choices? Choice overload in tourism. *Tourism Management*, *35*, 1–12. https://doi.org/10.1016/j.tourman.2012.05.004

Pekovic, S., & Rolland, S. (2020). Recipes for achieving customer loyalty: A qualitative comparative analysis of the dimensions of customer experience. *Journal of Retailing and Consumer Services*, *56*, 102171. https://doi.org/10.1016/j.jretconser.2020.102171

Petrescu, M. (2013). Marketing research using single-item indicators in structural equation models. *Journal of marketing analytics*, *1*(2), 99–117. https://doi.org/10.1057/jma.2013.7

Pieters, R., & Warlop, L. (1999). Visual attention during brand choice: The impact of time pressure and task motivation. *International Journal of Research in Marketing*, *16*(1), 1–16. https://doi.org/10.1016/s0167-8116(98)00022-6

Pocock, D. C. D. (1981). Sight and Knowledge. *Transactions of the Institute of British Geographers*, *6*(4), 385. https://doi.org/10.2307/621875

Pupil Labs. (2023). Pupil Labs v3.5.1 [Computer software]. Pupil Labs. Retrieved from https://www.pupil-labs.com/

Ranjbarian, B., Abdollahi, S. M., & Ghorbani, H. (2016). *Potential antecedents and consequences of online confusion in tourism industry*. https://doi.org/10.1109/ecdc.2016.7492984

Rayner, K., Li, X., Williams, C. C., Cave, K. R., & Well, A. D. (2007). Eye movements during information processing tasks: Individual differences and cultural effects. *Vision Research*, *47*(21), 2714–2726. https://doi.org/10.1016/j.visres.2007.05.007

Reichheld, F. F. (2003). The one number you need to grow. *Harvard business review*, *81*(12), 46-55.

Reichheld, F. F., Markey Jr, R. G., & Hopton, C. (2000). E-customer loyalty-applying the traditional rules of business for online success. *European business journal*, *12*(4), 173.

Rese, A., Ganster, L., & Baier, D. (2020). Chatbots in retailers' customer communication: How to measure their acceptance? *Journal of Retailing and Consumer Services*, *56*, 102176. https://doi.org/10.1016/j.jretconser.2020.102176

Riikkinen, M., Saarijärvi, H., Sarlin, P., & Lähteenmäki, I. (2018). Using artificial intelligence to create value in insurance. *International Journal of Bank Marketing*, *36*(6), 1145–1168. https://doi.org/10.1108/ijbm-01-2017-0015

Ringle, Christian M., Wende, Sven, & Becker, Jan-Michael. (2022). SmartPLS 4. Oststeinbek: SmartPLS. Retrieved from https://www.smartpls.com

Rowley, J. (2000). Product searching with shopping bots. Internet Research, *10*(3), 203–214. https://doi.org/10.1108/10662240010331957

Rybanská, J., Nagyová, Ľ., & Košičiarová, I. (2014). Sensory marketing: sensory perceptions and emotions as the key factors influencing consumer behaviour and decision-making. *Marketing Identity. University of Ss. Cyril and Methodius in Trnava, Trnava, Slovak Republic*.

Sainani, K. L. (2015). Dealing With Missing Data. *Pm&r*, *7*(9), 990–994. https://doi.org/10.1016/j.pmrj.2015.07.011

Salminen, J., Jansen, B. J., An, J., Jung, S., Nielsen, L. R., & Kwak, H. (2018). Fixation and Confusion. *Conference on Human Information Interaction and Retrieval*. https://doi.org/10.1145/3176349.3176391

Salminen, J., Nagpal, M., Kwak, H., An, J., Jung, S., & Jansen, B. J. (2019). Confusion Prediction from Eye-Tracking Data. *International Conference on Information Systems*. https://doi.org/10.1145/3361570.3361577

Salvucci, D. D., & Goldberg, J. F. (2000). *Identifying fixations and saccades in eye-tracking protocols*. https://doi.org/10.1145/355017.355028

Schmitt, B. (1999). Experiential Marketing. *Journal of Marketing Management*, *15*(1–3), 53–67. https://doi.org/10.1362/026725799784870496

Schütt, H. H., Rothkegel, L. O. M., Trukenbrod, H. A., Engbert, R., & Wichmann, F. A. (2019). Disentangling bottom-up versus top-down and low-level versus high-level influences on eye movements over time. *Journal of Vision*, *19*(3), 1. https://doi.org/10.1167/19.3.1

Sharma, A., Pandher, J. S., & Prakash, G. (2022). Consumer confusion and decision postponement in the online tourism domain: the moderating role of self-efficacy. *Journal of Hospitality and Tourism Insights*. https://doi.org/10.1108/jhti-03-2022-0096

Sheehan, B., Jin, H. M., & Gottlieb, U. (2020). Customer service chatbots: Anthropomorphism and adoption. *Journal of Business Research*, *115*, 14–24. https://doi.org/10.1016/j.jbusres.2020.04.030

Shiu, J. Y. (2015). Investigating consumer confusion in the retailing context: the causes and outcomes. *Total Quality Management & Business Excellence*, *28*(7–8), 746–764. https://doi.org/10.1080/14783363.2015.1121094

Shukla, P., Banerjee, M., & Adidam, P. T. (2010). Antecedents and Consequences of Consumer Confusion: Analysis of the Financial Services Industry. *Advances in Consumer Research*, *37*, 292–297. http://www.acrwebsite.org/volumes/v37/acr_v37_15120.pdf

Signals on the scales. (z.d.). LMU Munich. https://www.lmu.de/en/newsroom/news-overview/news/signals-on-the-scales.html

Steinfeld, A., Jenkins, O. C., & Scassellati, B. (2009). The oz of wizard. *Human-Robot Interaction*. https://doi.org/10.1145/1514095.1514115

Sproles, G. B., & Kendall, E. (1986). A Methodology for Profiling Consumers' Decision-Making Styles. *Journal of Consumer Affairs*, *20*(2), 267–279. https://doi.org/10.1111/j.1745-6606.1986.tb00382.x

Stupple, E. J. N, Ball, L. J. & Ellis, D. (2013). Matching bias in syllogistic reasoning: Evidence for a dual-process account from response times and confidence ratings, *Thinking & Reasoning*, 19:1, 54-77. https://doi.org/10.1080/13546783.2012.735622

Tay, S. W., Ryan, P., & Ryan, C. (2016). Systems 1 and 2 thinking processes and cognitive reflection testing in medical students. *Canadian medical education journal*, *7*(2), e97-103. https://doi.org/10.36834/cmej.36777

Teichert, M., & Bolz, J. (2018). How Senses Work Together: Cross-Modal Interactions between Primary Sensory Cortices. *Neural Plasticity*, *2018*, 1–11. https://doi.org/10.1155/2018/5380921

Tessier-Lavigne, M. (2000). Visual processing by the retina. *Principles of neural science*, 507-522.

Tjiptono, F., Arli, D., & Bucic, T. (2014). Consumer confusion proneness: insights from a developing economy. *Marketing Intelligence & Planning*, *32*(6), 722–734. https://doi.org/10.1108/mip-05-2013-0082

Torabi, A., Hamidi, H., & Safaie, N. (2021). Effect of Sensory Experience on Customer Word-of-mouth Intention, Considering the Roles of Customer Emotions, Satisfaction,

and Loyalty. (2021b). *International Journal of Engineering*, *34*(3), 682–699. https://doi.org/10.5829/ije.2021.34.03c.13

Turnbull, P. W., Leek, S., & Ying, G. (2000). Customer Confusion: The Mobile Phone Market. *Journal of Marketing Management*, *16*(1–3), 143–163. https://doi.org/10.1362/026725700785100523

Unurlu, Ç. (2019). The Mediating Role of Brand Performance on the Relationship between Confusion - Brand Loyalty and Uncertainty Avoidance - Brand Loyalty. *Dokuz Eylül Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, *34*(4), 491–510. https://doi.org/10.24988/ije.2019344879

Verhoef, P. C., Lemon, K. N., Parasuraman, A., Roggeveen, A., Tsiros, M., & Schlesinger, L. A. (2009). Customer Experience Creation: Determinants, Dynamics and Management Strategies. *Journal of Retailing*, *85*(1), 31–41. https://doi.org/10.1016/j.jretai.2008.11.001

Wachowiak, L., Tisnikar, P., Canal, G., Coles, A., Leonetti, M., & Celiktutan, O. (2022). Analysing Eye Gaze Patterns during Confusion and Errors in Human–Agent Collaborations. In *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. https://doi.org/10.1109/ro-man53752.2022.9900589

Waghmare, C. (2019). Design Messages in Chatbots. In: Introducing Azure Bot Service. Apress, Berkeley, CA. https://doi.org/10.1007/978-1-4842-4888-1_4

Walsh, G., Hennig-Thurau, T., & Mitchell, V. W. (2007). Consumer confusion proneness: scale development, validation, and application. *Journal of Marketing Management*, *23*(7–8), 697–721. https://doi.org/10.1362/026725707x230009

Walsh, G., & Mitchell, V. (2010). The effect of consumer confusion proneness on word of mouth, trust, and customer satisfaction. *European Journal of Marketing*, *44*(6), 838–859. https://doi.org/10.1108/03090561011032739

Wetzels, M., Odekerken-Schröder, G., & Van Oppen, C. (2009). Using PLS Path Modeling for Assessing Hierarchical Construct Models: Guidelines and Empirical Illustration. *Management Information Systems Quarterly*, *33*(1), 177. https://doi.org/10.2307/20650284

Whitler, K. A. (2014, 18 July). *Why Word Of Mouth Marketing Is The Most Important Social Media*. Forbes. https://www.forbes.com/sites/kimberlywhitler/2014/07/17/why-word-of-mouth-marketing-is-the-most-important-social-media/?sh=1fdc4d6954a8

Wu, C., & Liu, C. (2022). Refined use of the eye-mind hypothesis for scientific argumentation using multiple representations. *Instructional Science*, *50*(4), 551–569. https://doi.org/10.1007/s11251-022-09581-w

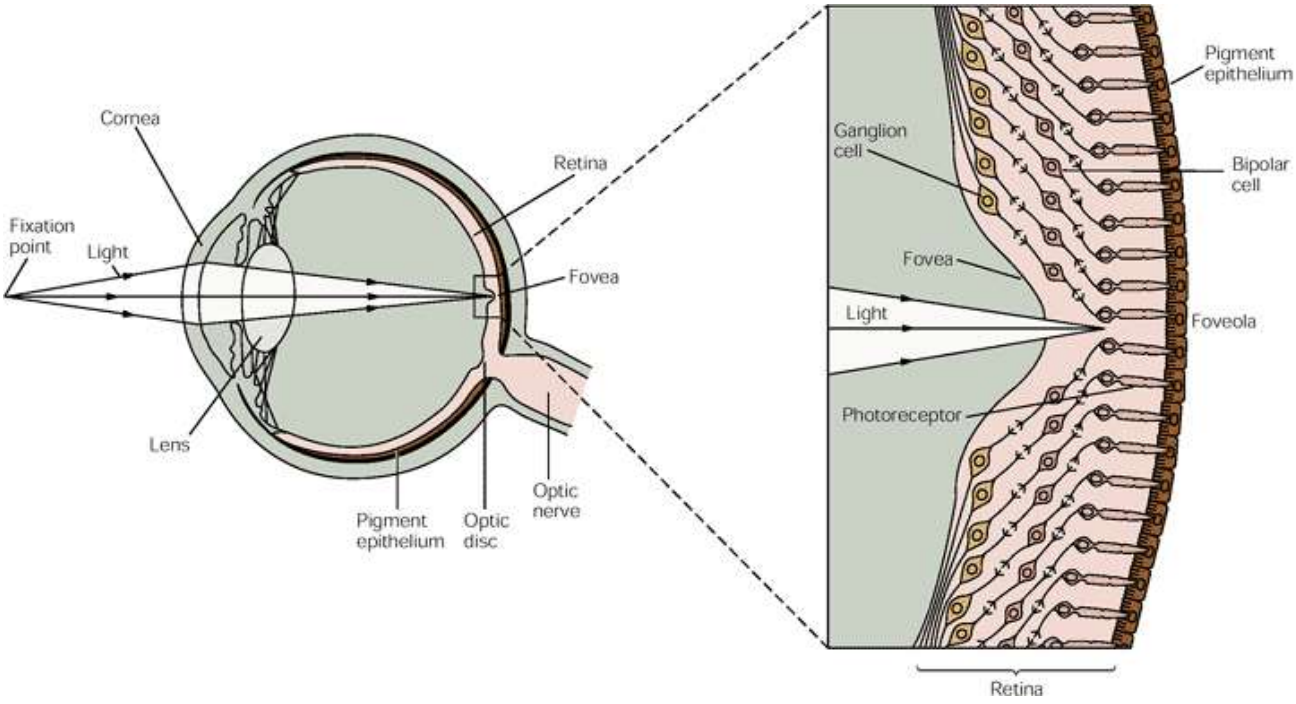
Zaid, S., & Patwayati, P. (2021). Impact of Customer Experience and Customer Engagement on Satisfaction and Loyalty: A Case Study in Indonesia. *Journal of Asian Finance, Economics and Business*, *8*(4), 983–992. https://doi.org/10.13106/jafeb.2021.vol8.no4.0983

Ziv, G. (2016). Gaze Behavior and Visual Attention: A Review of Eye Tracking Studies in Aviation. *The International Journal of Aviation Psychology*, *26*(3–4), 75–104. https://doi.org/10.1080/10508414.2017.1313096

Appendices

Appendix 1. Visualization of the human eye

Figure 3. The human eye



Appendix 2. Design of the experiment

Figure 1. First webpage of the experiment

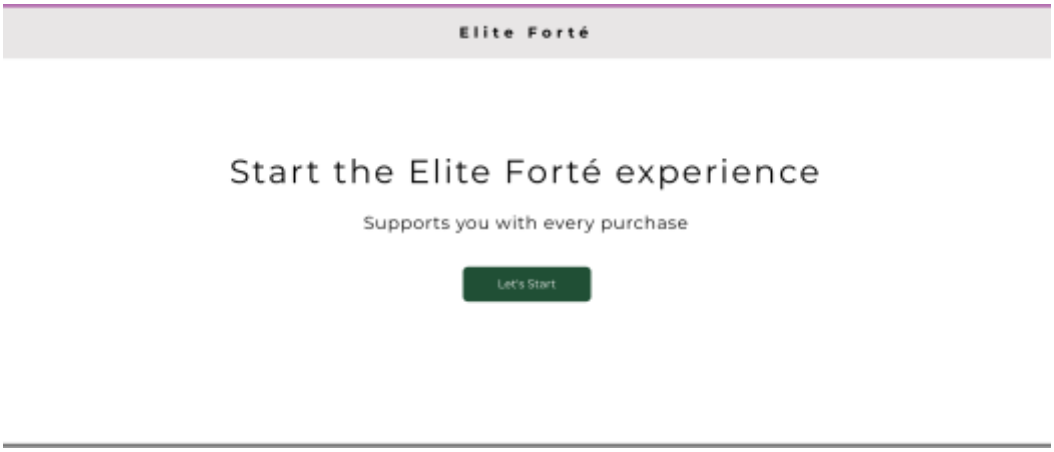


Figure 2. First webpage of the experiment

Figure 3. Error page condition 1

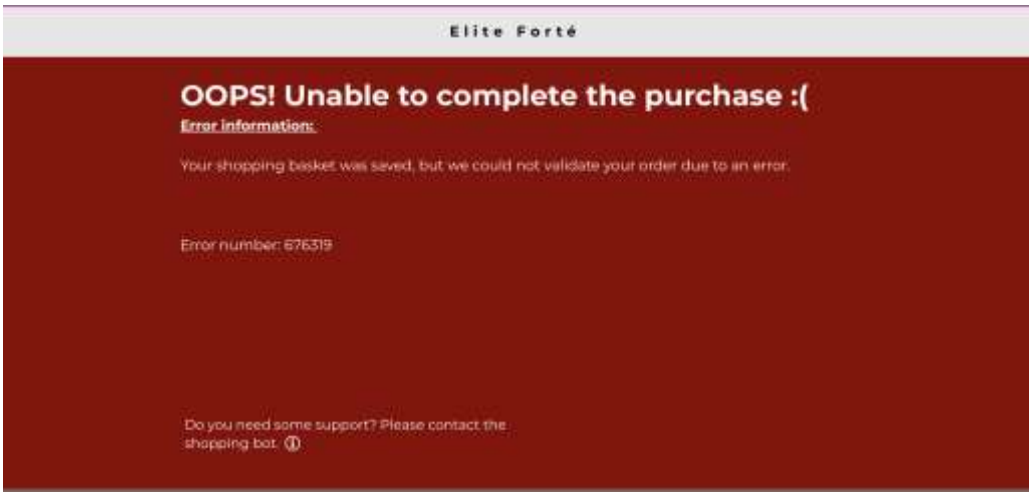


Figure 4. Error page condition 2

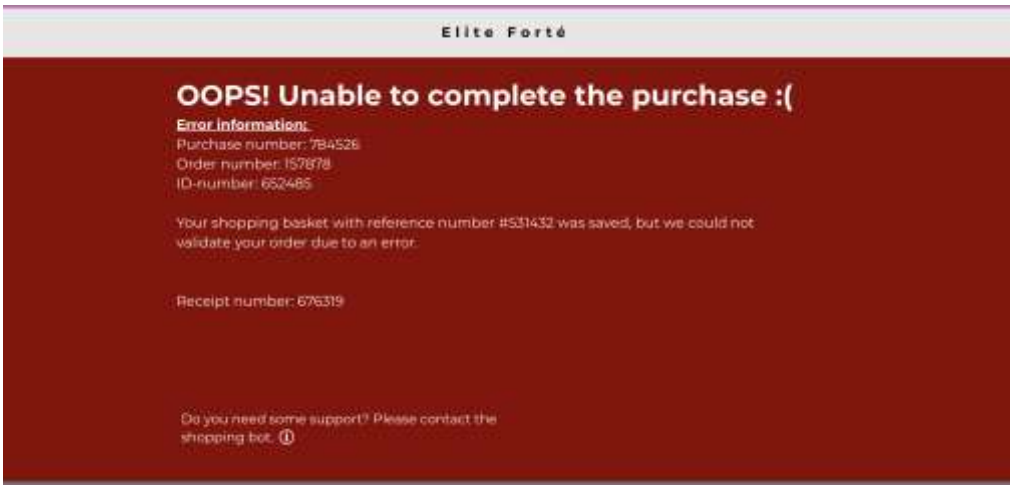


Figure 5. Error page condition 3

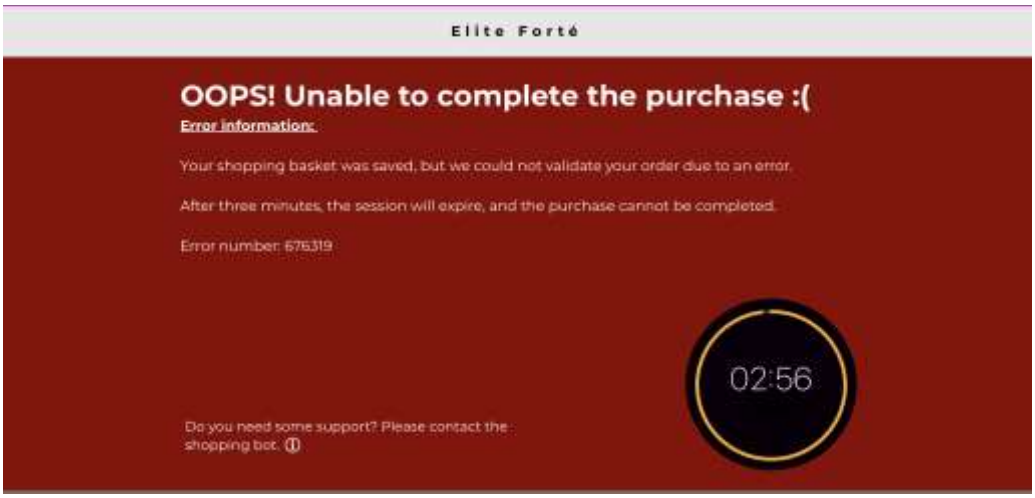


Figure 6. Error page condition 4

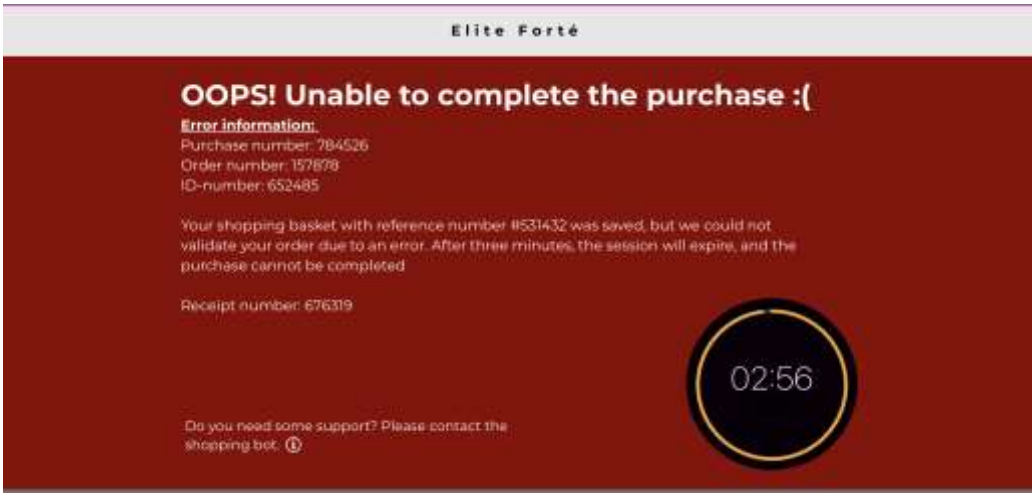
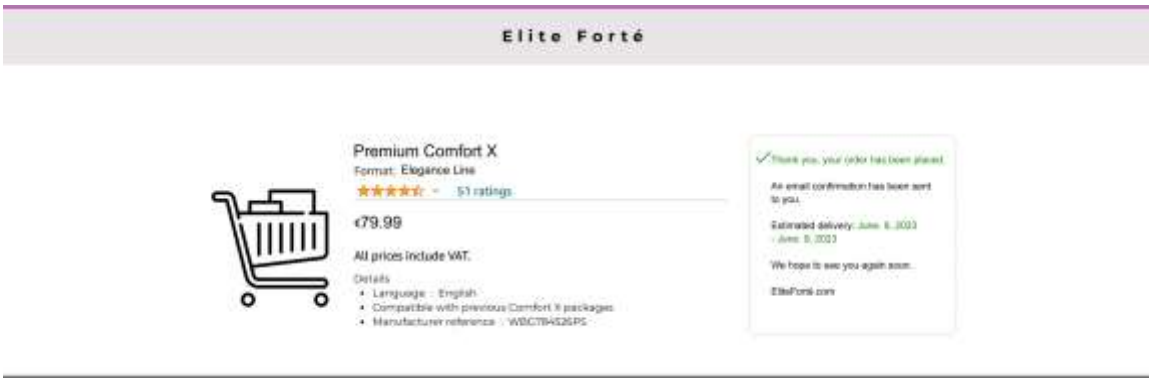


Figure 7. Final webpage of the experiment



Appendix 3. Manipulation check questions

Table 1. Questions manipulation check

Manipulation Check			
Confusion	- How strongly do you agree that the more information you sought, the harder it seemed to complete the purchase? - To what extent do you agree that the provided information was clear enough to complete the purchase? (-) - How strongly do you agree with: the information provided was so ambiguous that you often felt confused? - To what extent do you agree: that seeking more information made completing your purchase less confusing? (-)	7-point Likert scale; Adapted to chatbot context and rewrote to a more natural conversational text; All 4 items of the 'confused by overchoice consumer' factor are used.	Sproles and Kendall (1986)
Time pressure	- To what extent do you agree that you had enough time to complete the purchase? (-) - How strongly do you agree that you felt like you were under time pressure while making the purchase? - To what extent do you agree that the presence of a timer made you feel rushed during the purchase?	7-point Likert scale; Adapted to chatbot context and rewrote to a more natural conversational text; Based on the concept of perceived time risk.	De Dreu (2003)
Realness of chatbot	- To what extent do you agree that I behaved like an automated assistant?	7-point Likert scale	
Distraction	- To what extent do you agree that you sensed something in your surroundings that hindered you from fully concentrating on your experience with me?	7-point Likert scale	
Problem-solving preference	- To what extent do you agree that you would have preferred chatting to a human agent as opposed to me when encountering the error we had?	7-point Likert scale	

Appendix 4. Operationalization table

Table 1. Operationalization table

Construct	Operationalization	Measure	Source
CE-Sensations			
Dynamic gaze behavior on relevant areas	- Number of fixations - Average duration of fixations	Eye-tracking Eye-tracking	Meißner and Oll (2019)
CE-Behavior			
Searching for additional information	- Number of clicks on the information button - Total counts asking/seeking more information - Error page response time in seconds - Error message response time in seconds	Real time measure Real time measure Real time measure Real time measure	
CE-Outcomes			
Customer satisfaction	- To what extent would you say that you are satisfied with the shopping experience?	Seven-point Likert scale	Chung et al. (2020)
Customer loyalty	- How likely is it that you would use this shopping experience again? - How likely is it that you would tell a friend or colleague about this shopping experience?	Seven-point Likert scale Seven-point Likert scale	 Reichheld (2003)
Control variables	Operationalization	Measure	Source
Tolerance for ambiguity	- To what extent do you agree that you prefer situations in which there is some ambiguity? - How strongly would you agree that you enjoy tackling problems that are complex enough to be ambiguous? - To what extent do you agree that you generally	Seven-point Likert scale Seven-point Likert scale Seven-point Likert scale	McLain (1993)

	prefer novelty to familiarity?		
Confusion avoidance	- To what extent do you agree that providing clear information is the basis for avoiding confusion? - To what extent do you agree that you want to see unambiguous information when resolving confusing issues occurring in shopping experiences? - To what extent do you agree that you enjoy tackling issues pertaining to confusion occurring from ambiguous information?	Seven-point Likert scale Seven-point Likert scale Seven-point Likert scale	Schweizer, Kotouc, and Wagner (2006)
Credibility	- To what extent do you agree that I behaved like an automated assistant?	Seven-point Likert scale	
Distraction	- To what extent do you agree that you sensed something in your surroundings that hindered you from fully concentrating on your experience with me?	Seven-point Likert scale	
Problem solving preference	- To what extent do you agree that you would have preferred chatting to a human agent as opposed to me when encountering the error we had?	Seven-point Likert scale	
Age	- Can you please provide me with your age?		
Gender	- What is the gender you identify with the most?	(1) Male, (2) Female, (3) Other, (4) Prefer not to say	

Appendix 5. Full script

1. – Please interact with our Radboud shopping bot as you would normally in an online environment –

2. Hi, I am Cody, and will be your shopping assistant for the day. Before we begin, could you please provide me with your assigned participant ID?

Participants sends participant ID

3. Great, thank you! How are you doing today?

a. ***Answer with a positive or neutral expression:***

i. Great, let's continue then.

b. ***Answer with a negative expression:***

ii. That is unfortunate. I still hope you can help me out, let's continue.

c. ***Completely inappropriate answer: ***

iii. Okay, thank you for sharing, let's continue.

4. I am thankful that you are helping me today to improve my design. As a shopping bot my purpose is to enhance your shopping experience and help you complete a purchase. Let's start shopping, I will provide you with a link to a website where we can make the purchase.

5. https://shorturl.at/xyJN3

6. *** Delete the link when they have entered the site***

a. ***When asking why the link was deleted*:**

i. The link has expired, but you can still continue on the website.

b. ***When asking what to do***

iv. Please add the purchase to your basket.

7. *Wait for the error screen AND for the participant to contact shopping bot*

a. ***Not respond within 1 min of encountering the error***:

v. Oops something went wrong with adding the item to your basket ☹. Please send me the error number you see on the screen so I can try to solve the error for you.

b. *** Asks what they should do/ask for a number for the first time*:**

vi. It seems you stumbled upon an error when adding the item to your basket. Please send me the error number you see on the screen so I can try to solve the error for you.

c. *** Asks again for clarification etc. *:**

vii. I cannot provide further assistance without you providing me with the error number you see. Please send me the error number you see on the screen so I can try to solve the error for you.

8. ***3x when they send a number***

a. I am unable to use this number, please provide me with an alternative displayed error number so I can try to solve the error for you.

b. *** If the participant repeats the same number(s) *:**

i. You have already provided me with this number, but I am unable to use it. Please provide me with an alternative displayed error number so I can try to solve the error for you.

9. I think I managed to sort the problem out, can you please confirm to me if the purchase has been completed by checking this link https://shorturl.at/DFY47

10. * **Wait for the participant to confirm** *

a. *** If participant does not respond within 30 seconds *:**

viii. Is there something wrong? I am confident that I managed to fix the error for you, can you please check this link and confirm https://shorturl.at/DFY47

11. *** If participant confirms the purchase has been made *:**

a. Great to hear, it looks like we managed to resolve the issue together. To learn from this experience and help me improve for the next time, please help me in answering some questions. I would like you to reflect on your shopping experience with me. Please tell me how strongly you agree or disagree with my upcoming questions by typing in the corresponding numbers (when applicable). Let's get started.

12.

a. ***When answering in another way than a number*:**

ix. I was expecting a number, please reply to me with the respective number indicating your closest choice.

b. ***When participant answer off scale for example "8" *:**

x. Your choice is not reflected in the answer possibilities, please reply to me with the respective number indicating your closest choice.

c. ***When participant don't understand question***:

xi. I am afraid I am unable to provide you with further assistance. Please answer with the best of your abilities.

13. How strongly do you agree that the more information you sought, the harder it seemed to complete the purchase?

(1) Strongly Disagree (2) Disagree (3) Somewhat Disagree (4) Neither Disagree nor Agree (5) Somewhat Agree (6) Agree (7) Strongly Agree

14. Ok, and to what extent do you agree that the provided information was clear enough to complete the purchase?

(1) Strongly Disagree (2) Disagree (3) Somewhat Disagree (4) Neither Disagree nor Agree (5) Somewhat Agree (6) Agree (7) Strongly Agree

15. I see, and how strongly do you agree with: the information provided was so ambiguous that you often felt confused?

(1) Strongly Disagree (2) Disagree (3) Somewhat Disagree (4) Neither Disagree nor Agree (5) Somewhat Agree (6) Agree (7) Strongly Agree

16. Alright, and to what extent do you agree: that seeking more information made completing your purchase less confusing?

(1) Strongly Disagree (2) Disagree (3) Somewhat Disagree (4) Neither Disagree nor Agree (5) Somewhat Agree (6) Agree (7) Strongly Agree

17. Ok and moving on, to what extent do you agree that you had enough time to complete the purchase?

(1) Strongly Disagree (2) Disagree (3) Somewhat Disagree (4) Neither Disagree nor Agree (5) Somewhat Agree (6) Agree (7) Strongly Agree

18. Sounds good, and how strongly do you agree that you felt like you were under time pressure while making the purchase?

(1) Strongly Disagree (2) Disagree (3) Somewhat Disagree (4) Neither Disagree nor Agree (5) Somewhat Agree (6) Agree (7) Strongly Agree

19. Interesting… and to what extent do you agree that the presence of a timer made you feel rushed during the purchase?

(1) Strongly Disagree (2) Disagree (3) Somewhat Disagree (4) Neither Disagree nor Agree (5) Somewhat Agree (6) Agree (7) Strongly Agree

20. Ok, thank you for sharing. To help me further, please take me through what you were thinking as you were attempting to help me resolve the purchase error we encountered.

21. I see, and what actions crossed your mind in the process of resolving this error?

22. That is insightful thank you, and would you please elaborate by telling me what was valuable in achieving the outcome of the shopping experience and how?

23. I see, so to what extent do you agree that we have accomplished what we wanted from this purchase?

(1) Strongly Disagree (2) Disagree (3) Somewhat Disagree (4) Neither Disagree nor Agree (5) Somewhat Agree (6) Agree (7) Strongly Agree

24. Got it, and to what extent do you agree that we were effective in completing the purchase?

(1) Strongly Disagree (2) Disagree (3) Somewhat Disagree (4) Neither Disagree nor Agree (5) Somewhat Agree (6) Agree (7) Strongly Agree

25. Good, and to what extent do you agree that the way the error was resolved was useful?

(1) Strongly Disagree (2) Disagree (3) Somewhat Disagree (4) Neither Disagree nor Agree (5) Somewhat Agree (6) Agree (7) Strongly Agree

26. Moreover, how strongly do you agree that this shopping experience with me was valuable?

(1) Strongly Disagree (2) Disagree (3) Somewhat Disagree (4) Neither Disagree nor Agree (5) Somewhat Agree (6) Agree (7) Strongly Agree

27. Thank you for sharing, now in relation to how you felt, please tell me a bit about your feelings while attempting to help me resolve the purchase error.

28. Got it! And could you now please describe your overall feelings about the shopping experience outcome?

29. I see thank you for elaborating, so overall your shopping experience was …

(1 = Extremely Displeasing, 2 = Very Displeasing, 3 = Moderately Displeasing, 4 = Slightly Displeasing, 5 = Neutral, 6 = Slightly Nice, 7 = Moderately Nice, 8 = Very Nice, 9 = Extremely Nice)

30. Clear, and in the end, how pleasant was your shopping experience?

(1 = Extremely Unpleasant, 2 = Very Unpleasant, 3 = Moderately Unpleasant, 4 = Slightly Unpleasant, 5 = Neutral, 6 = Slightly Pleasant, 7 = Moderately Pleasant, 8 = Very Pleasant, 9 = Extremely Pleasant)

31. Got it, and how agreeable or disagreeable would you say your whole shopping experience was?

(1 = Extremely Disagreeable, 2 = Very Disagreeable, 3 = Moderately Disagreeable, 4 = Slightly Disagreeable, 5 = Neutral, 6 = Slightly Agreeable, 7 = Moderately Agreeable, 8 = Very Agreeable, 9 = Extremely Agreeable)

32. Thanks for that, so this shopping experience left you feeling...

(1 = Extremely Sad, 2 = Very Sad, 3 = Moderately Sad, 4 = Slightly Sad, 5 = Neutral, 6 =Slightly Happy, 7 = Moderately Happy, 8 = Very Happy, 9 = Extremely Happy)

33. Thanks for sharing how you felt with me. So as an outcome, to what extent would you say that you are satisfied with the shopping experience?

(1) Strongly Dissatisfied (2) Dissatisfied (3) Somewhat Dissatisfied (4) Neither Dissatisfied nor Satisfied (5) Somewhat Satisfied (6) Satisfied (7) Strongly Satisfied

34. Ok, and how likely is it that you would use this shopping experience again?

(1) Highly Unlikely (2) Unlikely (3) Somewhat Unlikely (4) Neither Unlikely nor Likely (5) Somewhat Likely (6) Likely (7) Strongly Likely

35. Thanks for sharing, and how likely is it that you would tell a friend or colleague about this shopping experience?

(1) Highly Unlikely (2) Unlikely (3) Somewhat Unlikely (4) Neither Unlikely nor Likely (5) Somewhat Likely (6) Likely (7) Strongly Likely

36. Thank you and moving on, to what extent do you agree that you prefer situations in which there is some ambiguity?

(1) Strongly Disagree (2) Disagree (3) Somewhat Disagree (4) Neither Disagree nor Agree (5) Somewhat Agree (6) Agree (7) Strongly Agree

37. Alright, and how strongly would you agree that you enjoy tackling problems that are complex enough to be ambiguous?

(1) Strongly Disagree (2) Disagree (3) Somewhat Disagree (4) Neither Disagree nor Agree (5) Somewhat Agree (6) Agree (7) Strongly Agree

38. Great, and to what extent do you agree that you generally prefer novelty to familiarity?

(1) Strongly Disagree (2) Disagree (3) Somewhat Disagree (4) Neither Disagree nor Agree (5) Somewhat Agree (6) Agree (7) Strongly Agree

39. Got it, and to what extent do you agree that providing clear information is the basis for avoiding confusion?

(1) Strongly Disagree (2) Disagree (3) Somewhat Disagree (4) Neither Disagree nor Agree (5) Somewhat Agree (6) Agree (7) Strongly Agree

40. Ok, and to what extent do you agree that you want to see unambiguous information when resolving confusing issues occurring in shopping experiences?

(1) Strongly Disagree (2) Disagree (3) Somewhat Disagree (4) Neither Disagree nor Agree (5) Somewhat Agree (6) Agree (7) Strongly Agree

41. Moreover, to what extent do you agree that you enjoy tackling issues pertaining to confusion occurring from ambiguous information?

(1) Strongly Disagree (2) Disagree (3) Somewhat Disagree (4) Neither Disagree nor Agree (5) Somewhat Agree (6) Agree (7) Strongly Agree

42. Almost done, a few questions more. To what extent do you agree that I behaved like an automated assistant?

(1) Strongly Disagree (2) Disagree (3) Somewhat Disagree (4) Neither Disagree nor Agree (5) Somewhat Agree (6) Agree (7) Strongly Agree

43. Got it, and to what extent do you agree that you sensed something in your surroundings that hindered you from fully concentrating on your experience with me?

(1) Strongly Disagree (2) Disagree (3) Somewhat Disagree (4) Neither Disagree nor Agree (5) Somewhat Agree (6) Agree (7) Strongly Agree

44. Alright, and to what extent do you agree that you would have preferred chatting to a human agent as opposed to me when encountering the error we had?

(1) Strongly Disagree (2) Disagree (3) Somewhat Disagree (4) Neither Disagree nor Agree (5) Somewhat Agree (6) Agree (7) Strongly Agree

45. Thank you so much for sharing this with me, a few last questions before you go. Compared to how you felt starting this experience with me, how would you describe feeling now? Can you elaborate on why you feel this way?

46. Thank you for elaborating, can you please provide me with your age?

47. Ok, before you go, what is the gender you identify with the most?
(1) Male, (2) Female, (3) Other, (4) Prefer not to say.

48. Thank you for participating and being patient with me! This is the end of the experience. Please do not close any windows, stay seated, and wait till someone helps you.

Appendix 6. Consent form

Consent form

Purpose:
The purpose of this study is to investigate responses to shopping bots.

Equipment:
Pupil Labs eye-tracking equipment, Empatica E4 wristband, iPhone XR and MacBook Pro 16-inch 2021.

Procedure:
During this experiment, you will be asked to interact with a shopping bot. Please confirm the following. I confirm that I do not have any physical, mental or health-related reasons or problems that should preclude my participation in this study, and I also confirm that I assume all of the physical, psychological, and financial risks associated with the use of the above indicated equipment.

If you agree to participate in this experiment, you will be asked to do the following:
Interact with our Radboud shopping bot as you would normally in an online environment, while wearing the Pupil Labs eye-tracking equipment and Empatica E4 wristband. The eye-tracking equipment measures your physiological response in terms of pupil fixations and pupil dilation during the experience, and the Empatica E4 wristband logs your level of arousal. Further, your facial expressions will be recorded using the iPhone XR, and other activities will be captured using screen recording software on a MacBook Pro 16-inch 2021. The total time that is required to complete this study is approximately 15 minutes including instructions, calibrating, and debriefing.

Health notice/risk
The devices used in this experiment are not expected to cause any physical harm or other discomforting feelings. I voluntarily assume all associated risks and take full responsibility for these and any other consequences that may arise from my participation.

Confidentiality:
Your participation is voluntary, and you can leave the lab experiment at any time, without a reason or permission, even after signing this consent form. The data that is provided by you will be kept confidential and stored in a responsible and correct way. In addition, these data are only used for research purposes and not for any other secondary purposes. Personal and sensitive data will not be recorded, and the experiment is completely anonymous. The information from this research might be published in the Radboud Thesis repository. Lastly, the information and results from this project may be submitted for publication in academic journals, however, this will not include personal data and data cannot be traced back to you in any way.

Contact:
If you are interested in the results of this experiment or have any questions, please contact the research team via Tao Zhu (s1083709, Master student, Business Administration Nijmegen School of Management), email:[1]

Statement of consent:
I have read the above information; I understand this completely and I consent to participate in this experiment.

_____
Name of Participant

_____ _____
Signature of Participant Date

Thank you for your participation!

[1] Email address is not shown due to privacy reasons

Appendix 7. Results manipulation check pre-test

Appendix 7A: Confirmatory factor analysis

Table 7A.1. KMO and Bartlett confusion

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.625
Bartlett's Test of Sphericity	Approx. Chi-Square	33.611
	df	6
	Sig.	<.001

Table 7A.2. Communalities confusion

Communalities

	Initial	Extraction
Confusion_item_1 The more information I sought, the harder it seems to complete my purchase.	1.000	.742
Reversed_Confusion_item_2 I find the provided information clear enough to complete the purchase.	1.000	.219
Confusion_item_3 The information provided to me was too ambiguous that often I feel confused.	1.000	.854
Reversed_confusion_item_4 All the information I sought, makes completing my purchase less confusing.	1.000	.735

Extraction Method: Principal Component Analysis.

Table 7A.3. Component matrix confusion

Component Matrix[a]

	Component 1
Confusion_item_1 The more information I sought, the harder it seems to complete my purchase.	.862
Reversed_Confusion_item_2 I find the provided information clear enough to complete the purchase.	.468
Confusion_item_3 The information provided to me was too ambiguous that often I feel confused.	.924
Reversed_confusion_item_4 All the information I sought, makes completing my purchase less confusing.	.857

Extraction Method: Principal Component Analysis.

a. 1 components extracted.

// **Deleted Reversed Confusion item 2**

Why is this done? This item has a low loading (<.70) and also the communality value is low.

Table 7A.4. KMO and Bartlett after deletion

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.726
Bartlett's Test of Sphericity	Approx. Chi-Square	27.811
	df	3
	Sig.	<.001

Table 7A.5. Communalities after deletion

Communalities

	Initial	Extraction
Confusion_item_1 The more information I sought, the harder it seems to complete my purchase.	1.000	.743
Confusion_item_3 The information provided to me was too ambiguous that often I feel confused.	1.000	.837
Reversed_confusion_item_4 All the information I sought, makes completing my purchase less confusing.	1.000	.827

Extraction Method: Principal Component Analysis.

Table 7A.6. Total variance explained

Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.407	80.224	80.224	2.407	80.224	80.224
2	.373	12.420	92.644			
3	.221	7.356	100.000			

Extraction Method: Principal Component Analysis.

Table 7A.7. Component matrix after deletion

Component Matrix[a]

	Component 1
Confusion_item_1 The more information I sought, the harder it seems to complete my purchase.	.862
Confusion_item_3 The information provided to me was too ambiguous that often I feel confused.	.915

Reversed_confusion_item_4 All the information I sought, makes completing my purchase less confusing.	.909

Extraction Method: Principal Component Analysis.
a. 1 components extracted.

Time pressure

Table 7A.8. KMO and Bartlett time pressure

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.519
Bartlett's Test of Sphericity	Approx. Chi-Square	26.084
	df	3
	Sig.	<.001

Table 7A.9. Communalities time pressure

Communalities

	Initial	Extraction
Reversed_Time_pressure_item_1 I felt like I had enough time to complete the purchase.	1.000	.627
Time_pressure_item_2 I felt I was under time pressure while making the purchase.	1.000	.909
Time_pressure_item_3 The presence of a timer made me feel rushed during the purchase.	1.000	.669

Extraction Method: Principal Component Analysis.

Table 7A.10. Total variance explained time pressure

Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.204	73.473	73.473	2.204	73.473	73.473
2	.641	21.363	94.837			
3	.155	5.163	100.000			

Extraction Method: Principal Component Analysis.

Table 7A.11. Component matrix time pressure

Component Matrix[a]

	Component 1
Reversed_Time_pressure_item_1 I felt like I had enough time to complete the purchase.	.792
Time_pressure_item_2 I felt I was under time pressure while making the purchase.	.953
Time_pressure_item_3 The presence of a timer made me feel rushed during the purchase.	.818

Extraction Method: Principal Component Analysis.
a. 1 components extracted.

// **No items had to be deleted**

Why not? All items loaded high enough on this component, the communalities were fine and KMO and Bartlett met the criteria.

// **Deleting the item of confusion showed an improvement for Cronbach's alpha**

Table 7A.12. Cronbach's alpha if item is deleted

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Confusion_item_1 The more information I sought, the harder it seems to complete my purchase.	11.85	19.713	.696	.689
Reversed_Confusion_item_2 I find the provided information clear enough to complete the purchase.	11.65	23.292	.321	.864
Confusion_item_3 The information provided to me was too ambiguous that often I feel confused.	11.90	14.305	.803	.609
Reversed_confusion_item_4 All the information I sought, makes completing my purchase less confusing.	12.30	21.063	.645	.718

// **Summated scales were created**

Table 7A.13. Cronbach's alpha confusion scale

Reliability Statistics

Cronbach's Alpha	N of Items
.864	3

Table 7A.14. Cronbach's alpha time pressure scale

Reliability Statistics

Cronbach's Alpha	N of Items
.811	3

Appendix 7B: Manipulation check pre-test with summated scales

Table 7B.1. Mean difference confusion

Group Statistics

	Confusion_Grouped No confusion: Condition 1 + 3, Confusion: Condition 2 + 4	N	Mean	Std. Deviation	Std. Error Mean
Manipulation_Confusion	1	10	2.6333333	.94868330	.30000000
	2	10	5.1333333	1.04468082	.33035708

Table 7B.2. Significance of mean difference confusion

	t	df	Significance One-Sided p	Significance Two-Sided p	Mean Difference	Std. Error Difference
Manipulation_Confusion	-5.602	18	<.001	<.001	-2.50000000	.44624635
	-5.602	17.835	<.001	<.001	-2.50000000	.44624635

Table 7B.3. Mean difference time pressure

Group Statistics

	PerceivedUrgency_Grouped No time pressure: Condition 1 + 2, Time pressure: Condition 3 + 4	N	Mean	Std. Deviation	Std. Error Mean
Manipulation_TimePressure	1	10	2.9666667	1.05934991	.33499585
	2	10	4.8666667	1.96387121	.62103060

Table 7B.4. Significance of mean difference time pressure

	t	df	Significance One-Sided p	Significance Two-Sided p	Mean Difference	Std. Error Difference
Manipulation_TimePressure	-2.693	18	.007	.015	-1.90000000	.70562117
	-2.693	13.829	.009	.018	-1.90000000	.70562117

Table 7B.5. Alternative explanations means confusion

Group Statistics

	Confusion_Grouped No confusion: Condition 1 + 3, Confusion: Condition 2 + 4	N	Mean	Std. Deviation	Std. Error Mean
Realness_item_1 The shopping bot behaved like an automated assistant.	1	10	5.60	1.265	.400
	2	10	5.90	.738	.233
Realness_item_2 The shopping bot provided me with information to complete my purchase.	1	10	4.90	2.183	.690
	2	10	4.80	1.476	.467

Table 7B.6. Alternative explanations mean difference significance confusion

	t	df	Significance One-Sided p	Significance Two-Sided p	Mean Difference	Std. Error Difference
Realness_item_1 The shopping bot behaved like an automated assistant.	-.648	18	.263	.525	-.300	.463
	-.648	14.489	.264	.527	-.300	.463
Realness_item_2 The shopping bot provided me with information to complete my purchase.	.120	18	.453	.906	.100	.833
	.120	15.804	.453	.906	.100	.833

Table 7B.7. Alternative explanations means time pressure

Group Statistics

	PerceivedUrgency_Grouped No time pressure: Condition 1 + 2, Time pressure: Condition 3 + 4	N	Mean	Std. Deviation	Std. Error Mean
Realness_item_1 The shopping bot behaved like an automated assistant.	1	10	5.70	1.160	.367
	2	10	5.80	.919	.291
	1	10	4.40	1.955	.618

Realness_item_2 The shopping bot provided me with information to complete my purchase.	2	10	5.30	1.636	.517

Table 7B.8. Alternative explanations mean difference significance time pressure

			Significance			
	t	df	One-Sided p	Two-Sided p	Mean Difference	Std. Error Difference
Realness_item_1 The shopping bot behaved like an automated assistant.	-.214	18	.417	.833	-.100	.468
	-.214	17.107	.417	.833	-.100	.468
Realness_item_2 The shopping bot provided me with information to complete my purchase.	-1.116	18	.139	.279	-.900	.806
	-1.116	17.459	.140	.279	-.900	.806

Appendix 8. Results manipulation check experiment

Appendix 8A: Confirmatory factor analysis

Table 8A.1. KMO and Bartlett

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.680
Bartlett's Test of Sphericity	Approx. Chi-Square	338.073
	df	21
	Sig.	<.001

Table 8A.2. Communalities

Communalities

	Initial	Extraction
MCConfusionQ1 MCConfusion Q1	1.000	.519
MCConfusionQ2 MCConfusion Q2	1.000	.651
MCConfusionQ3 MCConfusion Q3	1.000	.711
MCConfusionQ4 MCConfusion Q4	1.000	.424
MCUrgencyQ1 MCUrgency Q1	1.000	.439
MCUrgencyQ2 MCUrgency Q2	1.000	.767
MCUrgencyQ3 MCUrgency Q3	1.000	.706

Extraction Method: Principal Component Analysis.

Table 8A.3. Component matrix

Component Matrix[a]

	Component	
	1	2
MCConfusionQ1 MCConfusion Q1	.708	-.132
MCConfusionQ2 MCConfusion Q2	.739	-.324
MCConfusionQ3 MCConfusion Q3	.755	-.376

MCConfusionQ4 MCConfusion Q4	.466	-.455
MCUrgencyQ1 MCUrgency Q1	.465	.472
MCUrgencyQ2 MCUrgency Q2	.489	.726
MCUrgencyQ3 MCUrgency Q3	.328	.774

Extraction Method: Principal Component Analysis.
a. 2 components extracted.

// **Deleted MCConfusion Q4**

Why is this done? This item has a low loading on component 1 (<.70) and the communality value is not that high.

// **Deleted MCUrgency Q1**

Why is this done? This item has a low loading on component 2, and also loads on component 1, which indicates a cross loading. In addition, the communality value is not that high.

Table 8A.4. KMO and Bartlett after deletion

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.601
Bartlett's Test of Sphericity	Approx. Chi-Square	252.218
	df	10
	Sig.	<.001

Table 8A.5. Communalities after deletion

Communalities

	Initial	Extraction
MCConfusionQ1 MCConfusion Q1	1.000	.576
MCConfusionQ2 MCConfusion Q2	1.000	.703
MCConfusionQ3 MCConfusion Q3	1.000	.750
MCUrgencyQ2 MCUrgency Q2	1.000	.797
MCUrgencyQ3 MCUrgency Q3	1.000	.821

Extraction Method: Principal Component Analysis.

Table 8A.6. Total variance explained after deletion

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.095	41.908	41.908	2.095	41.908	41.908
2	1.551	31.016	72.924	1.551	31.016	72.924
3	.600	12.000	84.924			
4	.407	8.131	93.056			
5	.347	6.944	100.000			

Extraction Method: Principal Component Analysis.

Table 8A.7. Component matrix after deletion

Component Matrix[a]

	Component	
	1	2
MCConfusionQ1 MCConfusion Q1	.752	-.104
MCConfusionQ2 MCConfusion Q2	.740	-.393
MCConfusionQ3 MCConfusion Q3	.770	-.395
MCUrgencyQ2 MCUrgency Q2	.507	.735
MCUrgencyQ3 MCUrgency Q3	.363	.830

Extraction Method: Principal Component Analysis.
a. 2 components extracted.

// **Deleting the items showed good results for the improvement of Cronbach's alpha**

Table 8A.8. Cronbach's alpha confusion when item is deleted

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
MCConfusionQ1 MCConfusion Q1	12.19	20.287	.487	.694
MCConfusionQ2 MCConfusion Q2	12.72	16.585	.586	.636

MCConfusionQ3 MCConfusion Q3	11.97	16.454	.653	.591
MCConfusionQ4 MCConfusion Q4	12.98	22.490	.386	.743

Table 8A.9. Cronbach's alpha time pressure when item is deleted

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
MCUrgencyQ1 MCUrgency Q1	8.32	12.531	.400	.766
MCUrgencyQ2 MCUrgency Q2	7.04	7.852	.658	.450
MCUrgencyQ3 MCUrgency Q3	6.47	8.604	.571	.575

// **Summated scales were created for confusion and time pressure**

Table 8A.10. Reliability of confusion manipulation scale

Reliability Statistics

Cronbach's Alpha	N of Items
.743	3

Table 8A.11. Reliability of time pressure manipulations scale

Reliability Statistics

Cronbach's Alpha	N of Items
.769	2

Appendix 8B: Manipulation check experiment with summated scales

Table 8B.1. Mean difference confusion

Group Statistics

	Grouped_ Confusion	N	Mean	Std. Deviation	Std. Error Mean
Manipulation _Confusion	.00000	99	3.3400673	1.48230668	.14897743
	1.00000	100	5.3066667	.93765234	.09376523

Table 8B.2. Significance of mean difference confusion

	t	df	Significance One-Sided p	Significance Two-Sided p	Mean Difference	Std. Error Difference
Manipulation_Confusion	-11.196	197	<.001	<.001	-1.96659933	.17564846
	-11.172	165.338	<.001	<.001	-1.96659933	.17602895

Table 8B.3. Mean difference time pressure

Group Statistics

	Grouped_ TimePressure	N	Mean	Std. Deviation	Std. Error Mean
Manipulation_TimePressure	.00000	100	3.6250000	1.60393298	.16039330
	1.00000	100	4.6600000	1.80330785	.18033078

Table 8B.4. Significnace of mean difference time pressure

	t	df	Significance One-Sided p	Significance Two-Sided p	Mean Difference	Std. Error Difference
Manipulation_TimePressure	-4.289	198	<.001	<.001	-1.03500000	.24134043
	-4.289	195.343	<.001	<.001	-1.03500000	.24134043

Table 8B.5. Alternative explanations means of confusion

Group Statistics

	Grouped_ Confusion	N	Mean	Std. Deviation	Std. Error Mean
CredibilityQ1	.00000	98	5.77	1.217	.123
CredibilityQ1	1.00000	100	5.74	1.220	.122
DistractionQ1	.00000	100	2.67	1.718	.172
DistractionQ1	1.00000	100	2.35	1.591	.159

Problem-solving preferenceQ1	.00000	99	3.44	1.673	.168
	1.00000	100	4.52	1.761	.176
Problem-solving preferenceQ1					

Table 8B.6. Alternative explanations significance of difference means confusion

	t	df	Significance One-Sided p	Significance Two-Sided p	Mean Difference	Std. Error Difference
CredibilityQ1	.146	196	.442	.884	.025	.173
CredibilityQ1	.146	195.937	.442	.884	.025	.173
DistractionQ1	1.367	198	.087	.173	.320	.234
DistractionQ1	1.367	196.857	.087	.173	.320	.234
Problem-solving preferenceQ1	-4.416	197	<.001	<.001	-1.076	.244
Problem-solving preferenceQ1	-4.417	196.673	<.001	<.001	-1.076	.244

Table 8B.7. Alternative explanations means of time pressure

Group Statistics

	Grouped_TimePressure	N	Mean	Std. Deviation	Std. Error Mean
CredibilityQ1	.00000	100	5.71	1.094	.109
CredibilityQ1	1.00000	98	5.80	1.331	.134
DistractionQ1	.00000	100	2.30	1.508	.151
DistractionQ1	1.00000	100	2.72	1.781	.178
Problem-solving preferenceQ1	.00000	100	4.09	1.810	.181
Problem-solving preferenceQ1	1.00000	99	3.88	1.786	.179

Table 8B.8. Alternative explanations significance of difference means time pressure

	t	df	Significance One-Sided p	Significance Two-Sided p	Mean Difference	Std. Error Difference
CredibilityQ1	-.496	196	.310	.620	-.086	.173
CredibilityQ1	-.496	187.468	.310	.621	-.086	.173
DistractionQ1	-1.800	198	.037	.073	-.420	.233
DistractionQ1	-1.800	192.729	.037	.073	-.420	.233
	.829	197	.204	.408	.211	.255

Problem-solving preferenceQ1 Problem-solving preferenceQ1	.829	196.998	.204	.408	.211	.255

Appendix 9. Results

Appendix 9A: Data preparation

Table 9A.1. Missing data analysis

Univariate Statistics							
	N	Mean	Std. Deviation	Missing		No. of Extremes[a]	
				Count	Percent	Low	High
Participants	200	100.50	57.879	0	.0	0	0
Average duration	197	168.4338	24.22458	3	1.5	0	2
Average dispersion	197	1.2737	.08007	3	1.5	19	0
Total fixations S1 and S2	197	2143.701	1077.9820	3	1.5	0	1
Average duration S1 and S2	197	168.3357 8680203 0400	23.376840 231289012	3	1.5	0	4
Total fixations non-relevant areas	197	418.822	419.3992	3	1.5	0	13
Average duration fixations non-relevant areas	197	167.079	25.2329	3	1.5	0	1
Blink Rate	197	208.305	128.8502	3	1.5	0	6
Clicks	200	.33	.658	0	.0	.	.
Total counts of asking/seeking more information	200	1.77	1.388	0	.0	0	17
Error page response time in seconds	190	29.32105 2631577 217	14.054959 355472382	10	5.0	0	5
Error message response time in seconds	180	20.82777 7777777 904	10.923722 612027657	20	10.0	0	3
SatisfactionQ1	198	5.01	1.441	2	1.0	5	0
LoyaltyQ1	200	4.52	1.774	0	.0	0	0
LoyaltyQ2	200	4.48	1.692	0	.0	0	0
AmbiguityQ1	200	3.43	1.661	0	.0	0	0
AmbiguityQ2	200	4.73	1.395	0	.0	6	0
AmbiguityQ3	200	4.42	1.335	0	.0	0	0
Confusion AvoidanceQ1	200	6.65	.781	0	.0	5	0
Confusion AvoidanceQ2	200	5.86	1.534	0	.0	11	0

Confusion AvoidanceQ3	200	3.95	1.546	0	.0	0	0
CredibilityQ1	198	5.75	1.215	2	1.0	7	0
DistractionQ1	200	2.51	1.659	0	.0	0	0
Problem-solving preferenceQ1	199	3.98	1.796	1	.5	0	0
Age	199	25.15	8.132	1	.5	0	18
MCConfusionQ1	200	4.42	1.743	0	.0	0	0
MCConfusionQ2	199	3.90	2.091	1	.5	0	0
MCConfusionQ3	200	4.63	2.000	0	.0	0	0
MCConfusionQ4	200	3.63	1.605	0	.0	0	0
MCTimePressureQ1	199	2.59	1.498	1	.5	0	12
MCTimePressureQ2	200	3.86	1.977	0	.0	0	0
MCTimePressureQ3	200	4.43	1.971	0	.0	0	0
Gender	199			1	.5		

Table 9A.2. Little's MCAR test

	Chi-square	**DF**	**Sig.**
First missing value analysis	413,194	295	.000
Second missing value analysis	300,920	265	.064

Table 9A.3. Descriptive statistics of 'age'

Descriptive Statistics									
	N	Minimum	Maximum	Mean	Std. Deviation	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Age	199	18	61	25.15	8.132	2.804	.172	7.564	.343
Valid N (listwise)	199								

Appendix 9B: Evaluation of the measurement model

Figure 9B.1. Initial measurement model

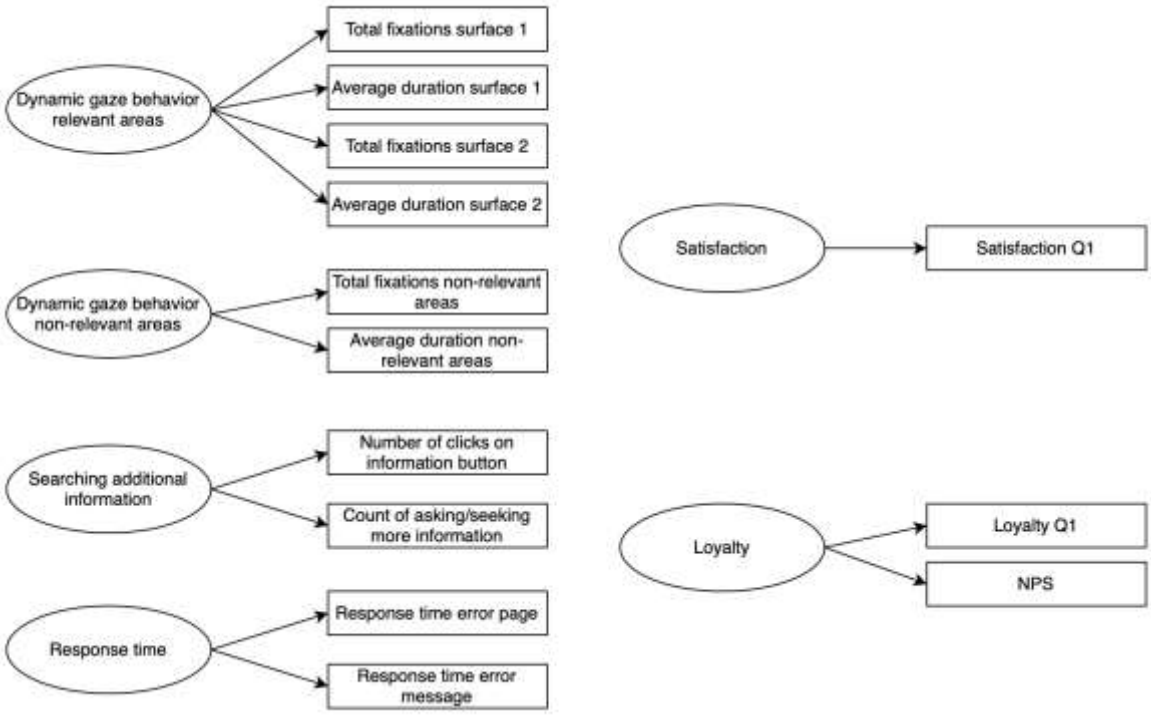


Figure 9B.2. Initial measurement model control variables

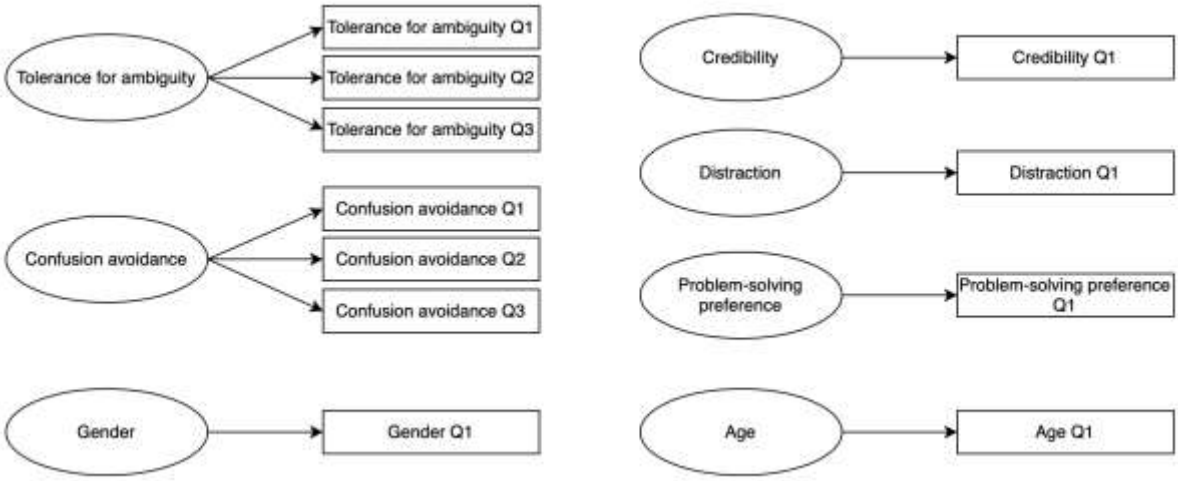


Table 9B.1. Factor loadings initial model

	Age	Credibility Q1	Distraction Q1	Problem-solving preference Q1	Confusion avoidance	Dummy YES Confusion
Age	1.000					
Credibility Q1		1.000				
Distraction Q1			1.000			
Problem-solving preference Q1				1.000		
Confusion avoidance Q1					-0.478	
Confusion avoidance Q2					0.409	
Confusion avoidance Q3					0.748	
Dummy YES Confusion						1.000

Table 9B.1. Continued

	Gender	Loyalty	Non-relevant areas	Relevant areas	Response time	Satisfaction	Searching additional information	Tolerance for ambiguity
Gender Q1	1.000							
LoyaltyQ1		0.907						
LoyaltyQ2		0.763						
Average duration non-relevant areas			0.895					
Total fixations non-relevant areas			0.577					
Average duration relevant areas				-0.112				
Total fixations relevant areas				0.939				
Error message response time in seconds					0.968			
Error page response time in seconds					0.376			
SatisfactionQ1						1.000		
Clicks							-0.070	
Total counts of asking/seeking more information							0.998	
Tolerance for ambiguity Q1								0.396
Tolerance for ambiguity Q2								-0.668
Tolerance for ambiguity Q3								0.149

Table 9B.2. Initial construct reliability and validity (convergent validity)

	Cronbach's alpha	Composite reliability (rho_a)	Composite reliability (rho_c)	Average variance extracted (AVE)
Confusion avoidance	0.083	-0.082	0.184	0.318
Loyalty	0.591	0.665	0.824	0.702
Non-relevant areas	0.265	0.322	0.715	0.567

Relevant areas	0.383	-0.897	0.382	0.447
Response time	0.235	0.499	0.663	0.540
Searching additional information	-0.287	-2.873	0.464	0.501
Tolerance for ambiguity	0.606	-1.251	0.006	0.208

// ‘**Relevant areas’ is split up into two separate variables**

Why is this done? The loading of average duration is negative and also very low and therefore does not measure the construct of relevant areas in a correct way (Asnawi et al, 2012). In addition, the construct reliability and validity statistics do not meet the criterion of >.70 and >.50 respectively. So, this construct is split into two unidimensional variables: total fixations relevant areas and average duration relevant areas.

// **‘Non-relevant areas’ is split up into two separate variables**

Why is this done? The loadings of the items are quite proper, although total fixation on non-relevant areas is a bit too low. However, the construct reliability statistics are also <.70, except rho c. In addition, due to the fact that the construct ‘relevant areas’ needs to be split up and needs to be compared to ‘non-relevant areas’, this construct is also split up into two unidimensional variables: total fixations non-relevant areas and average duration non-relevant areas.

Table 9B.3. Revised factor loadings after split up

	Age	**Credibility Q1**	**Distraction Q1**	**Problem-solving preference Q1**	**Average duration fixation non relevant areas**	**Average duration relevant areas**
Age_1	1.000					
Credibility Q1		1.000				
Distraction Q1			1.000			
Problem-solving preference Q1				1.000		
Average duration					1.000	

fixation non relevant areas						
Average duration S1 and S2						1.000

Table 9B.3. Continued

	Confusion Avoidance	**Dummy YES Confusion**	**Gender**	**Loyalty**	**Response time**	**Satisfaction**
Confusion AvoidanceQ1	-0.477					
Confusion AvoidanceQ2	0.409					
Confusion AvoidanceQ3	0.748					
Dummy YES Confusion		1.000				
Gender			1.000			
LoyaltyQ1				0.909		
LoyaltyQ2				0.760		
Error message response time in seconds					0.971	
Error page response time in seconds					0.368	
SatisfactionQ1						1.000

Table 9B.3. Continued

	Searching additional information	**Tolerance for ambiguity**	**Total fixation non relevant areas**	**Total fixations relevant areas**
Clicks	-0.075			
Total counts of asking/seeking more information	0.999			
AmbiguityQ1		0.393		
AmbiguityQ2		-0.671		
AmbiguityQ3		0.147		
Total fixations non relevant areas			1.000	
Total fixations S1 and S2				1.000

Table 9B.4. Revised construct reliability and validity (convergent validity) after split up

	Cronbach's alpha	**Composite reliability (rho_a)**	**Composite reliability (rho_c)**	**Average variance extracted (AVE)**
Confusion avoidance	0.083	-0.082	0.184	0.318
Loyalty	0.591	0.668	0.824	0.702
Response time	0.235	0.518	0.660	0.539
Searching additional information	-0.287	-2.502	0.461	0.502
Tolerance for ambiguity	0.606	-1.254	0.007	0.209

// **Deleted 'clicks'**

Why is this done? This item had a negative, very low factor loading on the construct of searching for additional information. In addition, the construct reliability statistics did not exceed the threshold (-.287 < .70; -2.502 < .70; .461 < .70). Although the convergence validity is just high enough (.502 > .50), the item clicks is still deleted due to the negative, low loading and the low values for construct validity. The construct 'searching for additional information' therefore becomes a unidimensional construct.

// **Deleted 'error page response time'**

Why is this done? This item has a loading of .368 which is below the minimal desired threshold of .70. In addition, the construct reliability statistics are lower than .70. Although the convergence validity is satisfactory (.539 > .50), the item is still deleted due to its low loading and the low construct reliability values. The construct 'response time' therefore becomes a unidimensional construct.

// **Deleted 'ambiguity Q2'**

Why is this done? This item has a negative loading on the construct tolerance for ambiguity which indicates that this item is not measuring it in the correct way (Asnawi et al, 2012). In addition, the construct reliability and validity values are below the threshold of 0.70 and 0.50 respectively.

// **Deleted 'confusion avoidance Q1'**

Why is this done? This item has a negative loading on the construct confusion avoidance which entails that this item does not measure this construct properly (Asnawi et al, 2012). Additionally, the values for construct reliability and validity are below the of 0.70 and 0.50 respectively.

// **Deleted 'confusion avoidance Q2'**

Why is this done? This item has a loading of .501 after deleting Q1 while Q3 has a loading of .866 after deleting Q1. Furthermore, after deleting Q1, the construct reliability values were still below the threshold of .70. Finally, that the construct of 'tolerance for ambiguity' and 'confusion avoidance' report a too high discriminant validity of 10.092 (>.9) and also 'confusion avoidance' has a high discriminant validity with other constructs in the model when two or three items are loading on it. This entails that confusion avoidance is related to other constructs, while this is not desirable (Hair et al. p. 775, 2019). When confusion avoidance is a one-item construct, the discriminant validity is below the threshold of .85.

Final measurement model

Table 9B.5. Final factor loadings

	Age	**Credibility Q1**	**Distraction Q1**	**Problem-solving preference Q1**	**Average duration fixation non relevant areas**	**Average duration relevant areas**
Age_1	1.000					
Credibility Q1		1.000				
Distraction Q1			1.000			
Problem-solving preference Q1				1.000		
Average duration fixation non relevant areas					1.000	
Average duration S1 and S2						1.000

Table 9B.5. Continued

	Confusion Avoidance	**Dummy YES Confusion**	**Gender**	**Loyalty**	**Response time**	**Satisfaction**
Confusion AvoidanceQ3	1.000					

Dummy YES Confusion		1.000				
Gender			1.000			
LoyaltyQ1				0.912		
LoyaltyQ2				0.755		
Error message response time in seconds					1.000	
SatisfactionQ1						1.000

Table 9B.5. Continued

	Searching additional information	**Tolerance for ambiguity**	**Total fixation non relevant areas**	**Total fixations relevant areas**
Total counts of asking/seeking more information	1.000			
AmbiguityQ1		0.942		
AmbiguityQ3		0.604		
Total fixation non relevant areas			1.000	
Total fixations S1 and S2				1.000

Table 9B.6. Final construct reliability and validity (convergent validity)

	Cronbach's alpha	**Composite reliability (rho_a)**	**Composite reliability (rho_c)**	**Average variance extracted (AVE)**
Loyalty	0.591	0.667	0.823	0.701
Tolerance for ambiguity	0.464	0.697	0.762	0.626

Table 9B.7. Discriminant validity (HTMT)

HTMT	Age	Alternative explainability & controlsQ	Alternative explainability & controlsQ	Alternative explainability & controlsQ	Average duration fixation non relevant areas	Average duration relevant areas	Confusion Avoidan	Dummy YES Confusi	Gender	Loyalty	Response time	Satisfa ction	Seeking additional informatic	Tolerance for ambiguit	Total fixation non relevant areas	Total fixations relevant areas
Age																
Alternative explainability & controlsQ1	0.133															
Alternative explainability & controlsQ2	0.055	0.061														
Alternative explainability & controlsQ3	0.195	0.078	0.121													
Average duration fixation non relevant areas	0.122	0.097	0.113	0.055												
Average duration relevant areas	0.199	0.099	0.094	0.025	0.532											
Confusion Avoidance	0.057	0.038	0.084	0.026	0.048	0.050										
Dummy YES Confusion	0.046	0.010	0.097	0.299	0.040	0.084	0.032									
Gender	0.060	0.037	0.028	0.075	0.154	0.102	0.025	0.004								
Loyalty	0.313	0.062	0.012	0.484	0.070	0.095	0.157	0.282	0.101							
Response time	0.029	0.059	0.011	0.100	0.095	0.005	0.010	0.530	0.075	0.070						
Satisfaction	0.247	0.067	0.031	0.409	0.041	0.095	0.111	0.318	0.044	0.847	0.105					
Seeking additional information	0.040	0.012	0.019	0.213	0.023	0.042	0.067	0.404	0.022	0.413	0.155	0.345				
Tolerance for ambiguity	0.023	0.103	0.071	0.073	0.108	0.135	0.494	0.074	0.166	0.187	0.076	0.061	0.146			
Total fixation non relevant areas	0.076	0.111	0.117	0.137	0.152	0.026	0.005	0.010	0.126	0.025	0.058	0.008	0.040	0.197		
Total fixations relevant areas	0.051	0.003	0.002	0.156	0.008	0.237	0.070	0.188	0.042	0.152	0.174	0.103	0.212	0.041	0.232	

Figure 9B.3. Final measurement model

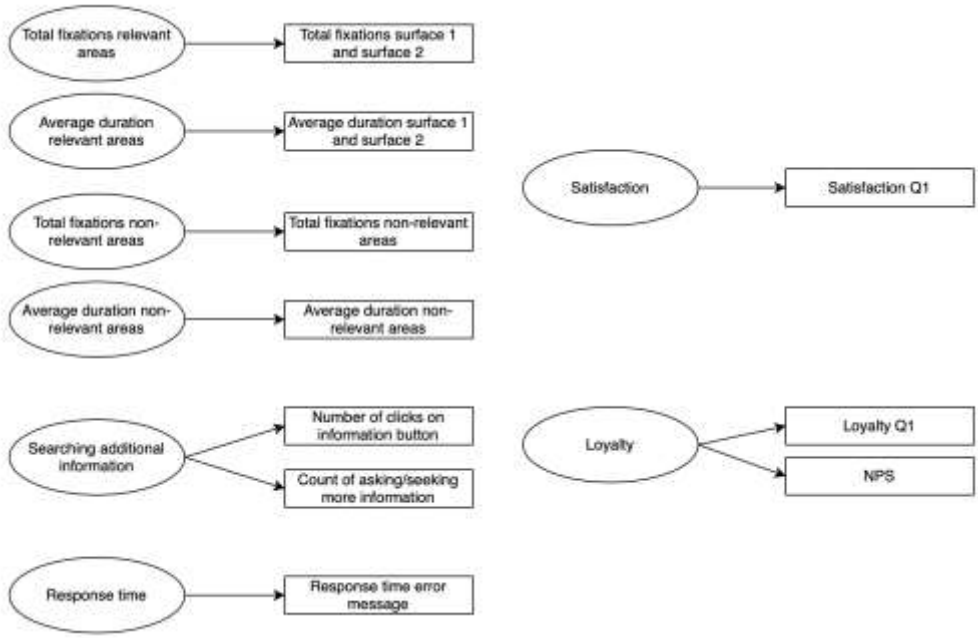


Figure 9B.4. Final measurement model control variables

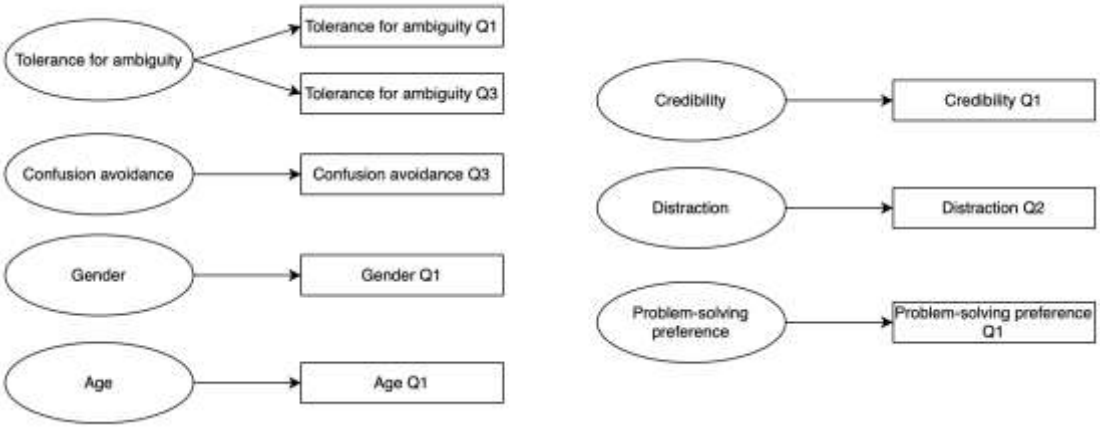


Table 9B.8. Factor loadings, composite reliability and average variance extracted of the constructs and their items.

Components and manifest variables	**Loading (t-value)**
Loyalty	**CR: 0.823, AVE: 0.701**
Q1 – How likely is it that you would use this shopping experience again?	0.912 (32.651) **
Q2 – How likely is it that you would tell a friend or colleague about this shopping experience?	0.755 (9.717) **
Tolerance for ambiguity	**CR: 0.726, AVE: 0.626**
Q1 – To what extent do you agree that you prefer situations in which there is some ambiguity?	0.942 (2.945) **
Q2 – To what extent do you agree that you generally prefer novelty to familiarity?	0.604 (1.602)
Notes: CR: Composite reliability; AVE: Average variance extracted, * p < .01	

Table 9B.9. GoF Index

GoF Index calculation	
Mean construct communality	Loyalty (0.912*0.912) + (0.755*0.755) = 0,7008845 Total fixations relevant = 1 Average duration relevant = 1 Average duration non-relevant = 1 Searching additional information = 1 Response time = 1 Satisfaction =1 **Mean construct communality:** 0.7008845 + 1 + 1 +1 +1 +1 +1 = 6.7008845 6.7008845 / 7 = 0,957269214
Mean R^2	0.014 + 0.017 + 0.231 + 0.279 + 0.263 + 0.170 + 0.033 = 1.007 1.007 / 7 = 0,143857143
Mean construct communality x mean R^2	0,957269214 * 0,143857143 = 0,137710014
GoF Index	√0,137710014 = 0,371092999276792

Appendix 9C: Evaluation of the structural model

Table 9C.1. VIF values among constructs

	VIF
Age_1	1.000
Credibility Q1	1.000
Distraction Q1	1.000
Problem-solving preference Q1	1.000
AmbiguityQ1	1.101
AmbiguityQ3	1.101
Average duration S1 and S2	1.000
Average duration fixation non relevant areas	1.000
Confusion AvoidanceQ3	1.000
Dummy YES Confusion	1.000
Dummy YES Pressure	1.000
Error message response time in seconds	1.000
Gender	1.000
LoyaltyQ1	1.214
LoyaltyQ2	1.214
SatisfactionQ1	1.000
Total counts of asking/seeking more information	1.000
Total fixation non relevant areas	1.000
Total fixations S1 and S2	1.000
Dummy YES TimePressure x Dummy YES Confusion	1.000

Table 9C.2. Adjusted R2 values of the constructs

	R-square	R-square adjusted
Average duration fixation non relevant areas	0.029	0.014
Average duration relevant areas	0.031	0.017
Loyalty	0.288	0.235
Response time	0.297	0.279
Satisfaction	0.316	0.264
Searching additional information	0.192	0.171
Total fixation non relevant areas	0.002	-0.014
Total fixations relevant areas	0.047	0.033

Table 9C.3. Adjusted R2 values of the constructs after deleting one constructs

	R-square	R-square adjusted
Average duration fixation non relevant areas	0.029	0.014
Average duration relevant areas	0.031	0.017
Loyalty	0.281	0.231
Response time	0.293	0.279
Satisfaction	0.311	0.263
Searching additional information	0.187	0.170
Total fixations relevant areas	0.047	0.033

Table 9C.4. Effect sizes (f2)

Effect sizes	Age	Alternative explainability & controlQ1	Alternative explainability & controlQ2	Alternative explainability & controlQ3	Duration non relevant areas	Duration relevant areas	Confusion Avoidance	Dummy YES Confusion	Dummy YES TimePressure	Gender	Loyalty	Response time	Satisfaction	Seeking additional information	Tolerance for ambiguity	Total fixations relevant areas	Dummy YES TimePressure x Dummy YES Confusion
Age											0.050		0.046				
Alternative explainability & controlQ1											0.005		0.007				
Alternative explainability & controlQ2											0.001		0.009				
Alternative explainability & controlQ3											0.094		0.108				
Average duration fixation non relevant areas											0.002	0.000	0.001	0.002			
Average duration relevant areas											0.002	0.001	0.001	0.004			
Confusion Avoidance											0.015		0.017				
Dummy YES Confusion					0.019	0.002					0.001	0.338	0.014	0.152		0.014	
Dummy YES TimePressure					0.024	0.029										0.009	
Gender											0.000		0.001				
Loyalty																	
Response time											0.001		0.002				
Satisfaction																	
Seeking additional information											0.073		0.065				
Tolerance for ambiguity											0.000		0.003				
Total fixations relevant areas											0.001	0.009	0.000	0.028			
Dummy YES TimePressure x Dummy YES Confusion					0.028	0.023										0.001	

Figure 9C.1. Simple Slope Analysis moderation effect on average duration relevant areas

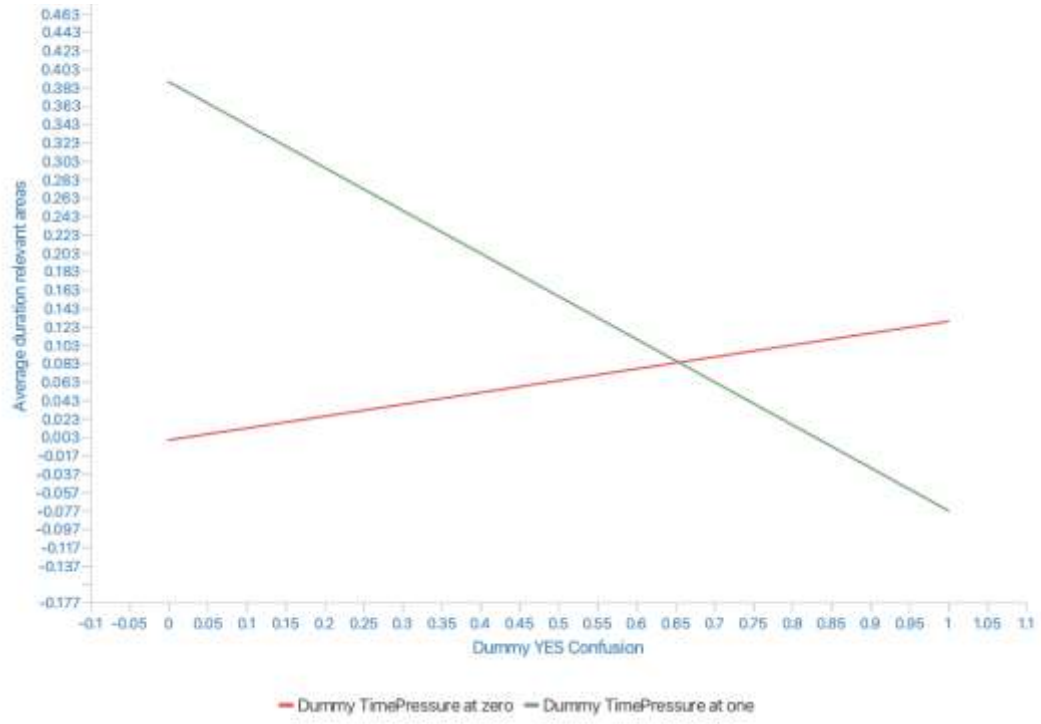


Table 9C.5. Serial mediation effects

Serial mediation effects	Original sample (O)	Sample mean (M)	Standard deviation (STDEV)	T statistics (|O/STDEV|)	P values
Dummy YES Confusion -> Total fixations relevant areas -> Searching additional information -> Satisfaction	-0.012	-0.012	0.011	1.130	0.129
Dummy YES Confusion -> Total fixations relevant areas -> Searching additional information -> Loyalty	-0.014	-0.014	0.014	0.997	0.159
Dummy YES Confusion -> Total fixations relevant areas -> Response time -> Satisfaction	0.001	0.001	0.004	0.281	0.389
Dummy YES Confusion -> Total fixations relevant areas -> Response time -> Loyalty	0.001	0.001	0.004	0.274	0.392
Dummy YES Confusion -> Average duration relevant areas -> Searching additional information -> Satisfaction	0.002	0.002	0.006	0.398	0.345
Dummy YES Confusion -> Average duration relevant areas -> Searching additional information -> Loyalty	0.003	0.003	0.006	0.399	0.345
Dummy YES Confusion -> Average duration relevant areas -> Response time -> Satisfaction	-0.000	-0.000	0.002	0.089	0.465
Dummy YES Confusion -> Average duration relevant areas -> Response time -> Loyalty	-0.000	-0.000	0.002	0.084	0.467

Table 9C.6. Serial mediation effects with moderation

Serial mediation effects with moderation	Original sample (O)	Sample mean (M)	Standard deviation (STDEV)	T statistics (|O/STDEV|)	P values
Dummy TimePressure x Dummy YES Confusion -> Total fixations relevant areas -> Searching additional information -> Satisfaction	-0.004	-0.005	0.012	0.324	0.373
Dummy TimePressure x Dummy YES Confusion -> Total fixations relevant areas -> Searching additional information -> Loyalty	-0.004	-0.006	0.015	0.293	0.385
Dummy TimePressure x Dummy YES Confusion -> Total fixations relevant areas -> Response time -> Satisfaction	0.000	0.000	0.003	0.114	0.455
Dummy TimePressure x Dummy YES Confusion -> Total fixations relevant areas -> Response time -> Loyalty	0.000	0.000	0.003	0.117	0.454
Dummy TimePressure x Dummy YES Confusion -> Average duration relevant areas -> Searching additional information -> Satisfaction	-0.011	-0.011	0.013	0.791	0.215
Dummy TimePressure x Dummy YES Confusion -> Average duration relevant areas -> Searching additional information -> Loyalty	-0.012	-0.012	0.015	0.791	0.215
Dummy TimePressure x Dummy YES Confusion -> Average duration relevant areas -> Response time -> Satisfaction	0.001	0.001	0.005	0.145	0.442
Dummy TimePressure x Dummy YES Confusion -> Average duration relevant areas -> Response time -> Loyalty	0.001	0.001	0.005	0.139	0.445

Table 9C.7. Mediation effects sensations

Mediation effects confusion > Sensations > CE-Outcomes	Original sample (O)	Sample mean (M)	Standard deviation (STDEV)	T statistics (|O/STDEV|)	P values
Dummy YES Confusion -> Total fixations relevant areas -> Satisfaction	0.002	0.003	0.025	0.074	0.470
Dummy YES Confusion -> Total fixations relevant areas -> Loyalty	-0.008	-0.006	0.029	0.290	0.386
Dummy YES Confusion -> Average duration relevant areas -> Satisfaction	0.005	0.005	0.019	0.258	0.398
Dummy YES Confusion -> Average duration relevant areas -> Loyalty	0.006	0.009	0.023	0.280	0.390

Table 9C.8. Mediation effects sensations with moderation

Mediation effects confusion x perceived time pressure > Sensations > CE-Outcomes	Original sample (O)	Sample mean (M)	Standard deviation (STDEV)	T statistics (|O/STDEV|)	P values
Dummy TimePressure x Dummy YES Confusion -> Total fixations relevant areas -> Satisfaction	0.001	0.002	0.020	0.031	0.488
Dummy TimePressure x Dummy YES Confusion -> Total fixations relevant areas -> Loyalty	-0.003	-0.004	0.024	0.116	0.454

Dummy TimePressure x Dummy YES Confusion -> Average duration relevant areas -> Satisfaction	-0.022	-0.021	0.050	0.448	0.327
Dummy TimePressure x Dummy YES Confusion -> Average duration relevant areas -> Loyalty	-0.029	-0.030	0.058	0.502	0.308

Table 9C.9. Mediation effects behavior

Mediation effects confusion > Behavior > CE-Outcomes	**Original sample (O)**	**Sample mean (M)**	**Standard deviation (STDEV)**	**T statistics (|O/STDEV|)**	**P values**
Dummy YES Confusion -> Searching additional information -> Satisfaction	-0.172	-0.176	0.056	3.073	0.001
Dummy YES Confusion -> Searching additional information -> Loyalty	-0.190	-0.185	0.057	3.332	0.000
Dummy YES Confusion -> Response time -> Satisfaction	0.040	0.043	0.082	0.489	0.312
Dummy YES Confusion -> Response time -> Loyalty	0.037	0.033	0.078	0.474	0.318

Figure 9B.2. Simple Slope Analysis moderation effect on average duration non-relevant areas

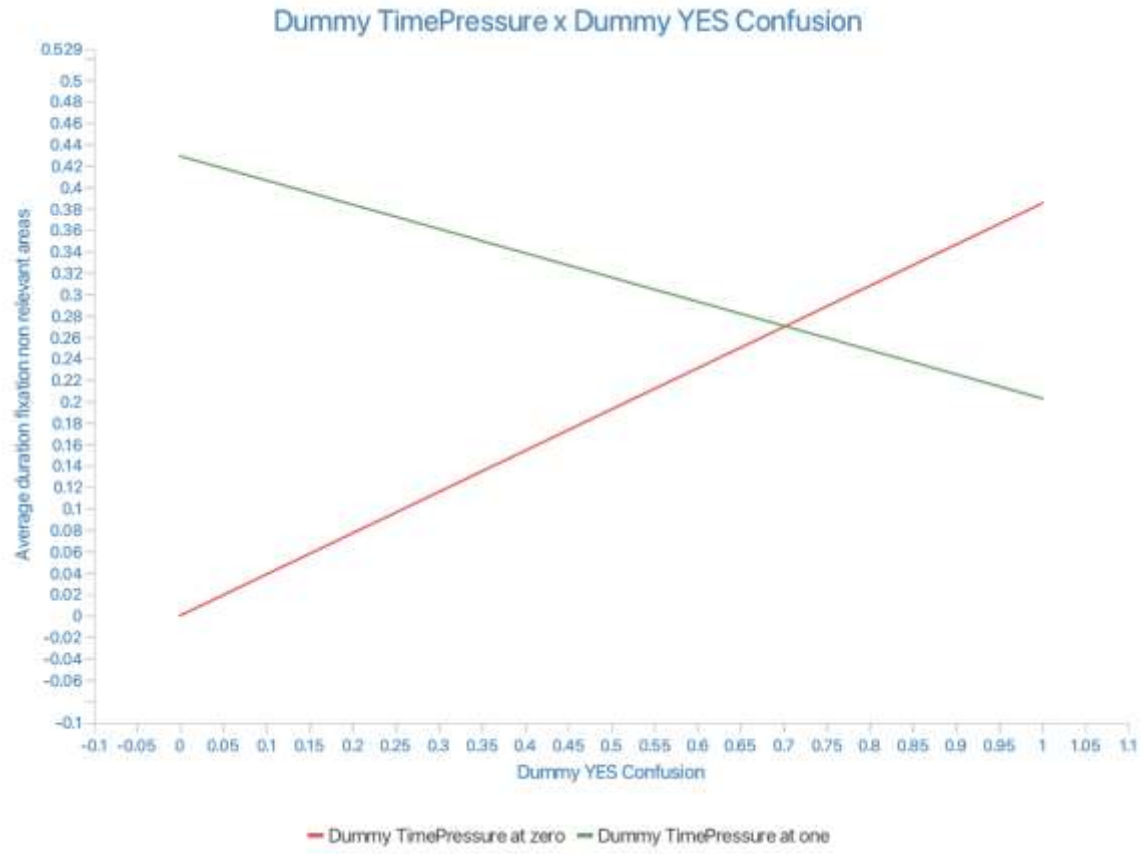


Table 9C.10. Serial mediation effects additional analysis

Serial mediation effects additional analysis	Original sample (O)	Sample mean (M)	Standard deviation (STDEV)	T statistics (|O/STDEV|)	P values
Dummy YES Confusion -> Average duration fixation non relevant areas -> Searching additional information -> Satisfaction	-0.005	-0.005	0.008	0.536	0.296
Dummy YES Confusion -> Average duration fixation non relevant areas -> Searching additional information -> Loyalty	-0.005	-0.005	0.009	0.580	0.281
Dummy YES Confusion -> Average duration fixation non relevant areas ->	0.001	0.002	0.005	0.296	0.384

Response time -> Satisfaction					
Dummy YES Confusion -> Average duration fixation non relevant areas -> Response time -> Loyalty	0.001	0.002	0.005	0.277	0.391

Table 9C.11. Serial mediation effects with moderation additional analysis

Serial mediation effects with moderation additional analysis	Original sample (O)	Sample mean (M)	Standard deviation (STDEV)	T statistics (|O/STDEV|)	P values
Dummy TimePressure x Dummy YES Confusion -> Average duration fixation non relevant areas -> Searching additional information -> Satisfaction	0.007	0.008	0.014	0.530	0.298
Dummy TimePressure x Dummy YES Confusion -> Average duration fixation non relevant areas -> Searching additional information -> Loyalty	0.008	0.009	0.014	0.570	0.284
Dummy TimePressure x Dummy YES Confusion -> Average duration fixation non relevant areas -> Response time -> Satisfaction	-0.002	-0.003	0.007	0.327	0.372
Dummy TimePressure x Dummy YES Confusion -> Average duration fixation non relevant areas -> Response time -> Loyalty	-0.002	-0.002	0.007	0.314	0.377

Table 9C.12. Mediation effects with and without moderation additional analysis

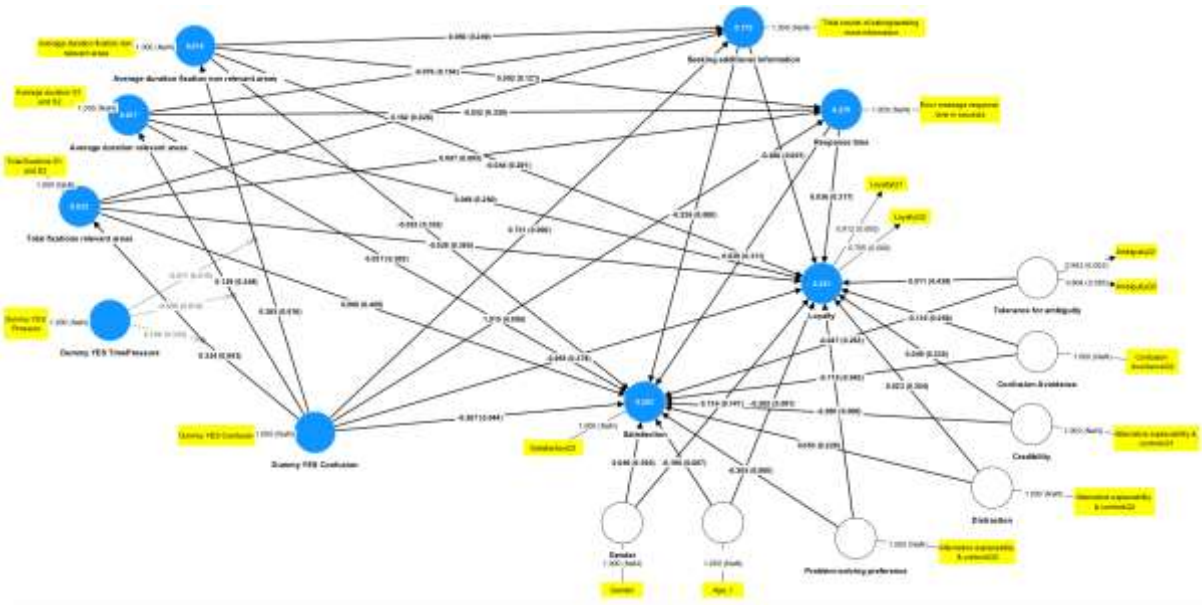
Mediation effects additional analysis, with and without moderation	Original sample (O)	Sample mean (M)	Standard deviation (STDEV)	T statistics (|O/STDEV|)	P values
Dummy YES Confusion -> Average duration fixation non relevant areas -> Satisfaction	-0.013	-0.011	0.032	0.386	0.350
Dummy YES Confusion -> Average duration fixation non relevant areas -> Loyalty	-0.017	-0.018	0.037	0.468	0.320
Dummy TimePressure x Dummy YES Confusion -> Average duration fixation non relevant areas -> Satisfaction	0.020	0.015	0.050	0.398	0.345
Dummy TimePressure x Dummy YES Confusion -> Average duration fixation non relevant areas -> Loyalty	0.027	0.026	0.057	0.479	0.316

Table 9C.13. Control variables

Control variables	Original sample (O)	Sample mean (M)	Standard deviation (STDEV)	T statistics (|O/STDEV|)	P values
Age -> Loyalty	-0.202	-0.203	0.064	3.171	0.001
Age -> Satisfaction	-0.190	-0.190	0.077	2.474	0.007
Confusion Avoidance -> Loyalty	-0.110	-0.105	0.070	1.559	0.060
Confusion Avoidance -> Satisfaction	-0.116	-0.103	0.067	1.733	0.042
Credibility -> Loyalty	0.049	0.043	0.069	0.710	0.239

Credibility -> Satisfaction	0.070	0.066	0.081	0.871	0.192
Distraction -> Loyalty	0.023	0.027	0.063	0.372	0.355
Distraction -> Satisfaction	0.053	0.054	0.072	0.746	0.228
Gender -> Loyalty	0.134	0.132	0.128	1.048	0.147
Gender -> Satisfaction	0.049	0.041	0.126	0.385	0.350
Problem-solving preference -> Loyalty	-0.290	-0.288	0.073	3.973	0.000
Problem-solving preference -> Satisfaction	-0.305	-0.300	0.064	4.739	0.000
Tolerance for ambiguity -> Loyalty	0.011	0.030	0.074	0.153	0.439
Tolerance for ambiguity -> Satisfaction	-0.047	-0.021	0.074	0.637	0.262

Figure 9C.3. Visual representation whole model in SmartPLS



Appendix 9D: Additional qualitative analysis behavior

I see, and what actions crossed your mind in the process of resolving this error?	
Answers of participants	**Interpretation**
I tried to ask you if you could help me but you didn't have the right answer.	Asking for help
It was a bit frustrating that you could not do anything with 4 of the 5 numbers I have send you, therefore I considered to not buy the product	Considering not to buy the product
I tried out the different numbers	Sending all the information available
I wanted to quit	Quitting with the purchase
asking the chat bot about where to find it	Asking for help
I expected the men who started this test would come and help me	
what do you mean with errornumber, and what information do you want	Searching additional information
just copy paste all numbers, or asking you which number you meant	Asking for information
Stopping the entire process of purchasing	Quit the purchase process
It was quite clear on the page that to resolve the error I would need to contact the shopping bot, so that was quite intuitive. The 'actions' I had to choose were mostly what information would be relevant to the bot	Intuitive actions
to find out which number I had to enter	Searching additional information
To try all the numbers	Sending all the information available
I just entered all numbers on the screen until it was the right one	
I thought that I did not understand the experiment correctly and thought about asking the researchers for help	Asking for help
Try and ask you first, then just trying the different numbers	Asking for help
To ask the bot for help, otherwise I would have tried to refresh the browser/restart the page.	Asking for help, refresh the page
Mentioning every single number to you untill I got the right one	Sending all the information available
I thought about leaving because I did not know how to make you understand what error I got	Leaving/quitting the purchase process

Well, when you didn't answer my question, I decided to send you each number I saw on the screen. I didn't see another option	Sending all the information available
Listing off all the availible numbers until my problem was resolved	Sending all the information available
searching for the right number	Searching for information
To just try to input every number I saw, starting with the number I thought was most likely to be the correct one.	Sending all the information available
Just try every number that is referred to, and see which one would work. I did think about which could would be most logic. But in the end I had to try 3 times before using the right code	Logical thinking, sending all the information available
trial and error by typing in every code I saw on the screen	Sending all the information available
I saw a lot of numbers I had to give you. For me, it was not really clear which number I had to give you since there were multiple options but none matched your description literally	Sending all the information available
First I tried the most logical number for me, because that was related to the error. After that I looked for a different code but couldn't really find it, so I just started to go from the top to the bottom to find out which one was the correct one	Logical thinking, sending all the information available
frustration. Looking for an option to seek help	Seeking help
To refresh the page. Or to ask you what code you wanted to use. I also thought that maybe I should get help outside the chat bot, like a customer service or something a person.	Refreshing the page
To do it by myself and trying different numbers before asking you	Sending all the information available
asking you were i could find the number you were looking for of quiting	Asking for more information
I was thinking why there was no specific message with the error code	
i asked youbut you could not help meso i tried every number	Asking for help
first i asked but that did not workafter that i just tried all the numbers	Asking for help, sending all the information available
Call help support. This worked	Asking for help
Finding the right number out of all these numbers	Looking for the right information
I like techs and computers so I found it interesting	

why did you even make me go through this	
I thought I would try the receipt number first, because it was center screen, but I really was not sure	Intuitive action (saw a number on prominent place)
To ask for help	Asking for help/more information
Could you elaborate on what you mean with this question?I am sorry I am not programmed to offer further assistance. I will however take this as feedback to improve my abilities in the future. Please continue with you the best of your ability.Well, the actions that crossed my mind were to first try and find out which number you wanted. After you failed to be clear about this i decided to try all the numbers given on the screen	
giving all the numeric information present	Sending all the information available (intuitive)
I was struggling to find the correct number in order to solve the problem	Not finding the right information
just providing all the numbers that were on the screen of the purchase error	Sending all the information available
that the process is annoying	
I looked more often to see if there was an error number i missed, so then i tried the different numbers based on likelihood that that is the one you meant	Searching for information, logical thinking on what number to send
I tried matching name of the number you needed with anything on the shopping page. I could not match anything. Then I tried asking you. But you did not answer. Then I trialled and errored	Asking for help, sending all the information available
Just throwing all the number i saw, one by one. And finally, you solved the issue.	Sending all the information available
Explaining there was no error number, asking which number you needed that was on the screening, entering the numbers one by one	Asking for more information
what i could do wrong	
In real life if you have multiple windows open how would you know which site we need help with?	
I tried to send multiple numbers that were on the announcement	Sending all the information available

Asking you to resolve the error and which number I could use to reply on your question	Asking for help and more information
mentioning them all, just to be safe and give 1 message instead of more with other numbers, that might have made it more complex	Sending all the information available
to type all numbers that were mentioned in the screen and then one of them would be the right one	Sending all the information available
listing all the visible numbers on the screen	Sending all the information available
I tried to get clarification from you, but you did not help me any further. So I just decided to name all the possibilities instead	Asking for help/more information, sending all the information available
That I wanted to solve it quickly	Urge to solve confusion
I wanted to get help from you	Asking/seeking help
just providing the different kind of numbers	Sending all the information available
send all the numbers, because your reaction time was very fast	Sending all the information available
How i could find the correct number, which would help to resolve the problem. I started using most of the numbers that were visible on my screen.	Asking for help/more information, sending all the information available
Talking to you	Asking for help
I tried to ask for more information and I wanted to just try all of the numbers that were shown in the screen	Asking for more information, sending all the information available
i did not know what you meant with the error number because there were 4/5 numbers. so i tryed every number.	Sending all the information available
Not sure what you mean by actions. I was considering not buying the product I suppose and trying a different supplier.	Considering to not buy the product
I thought about giving up and starting again, but since you deleted the first link, that wasn't a possibility anymore	Giving up and starting again
It cann't be difficult because you have 3 opportunities to give for resolving the error	
to give you all the numbers as fast as possible so at some point you would have the right one	Sending all the information available
restarting the purchasing attempt	Restarting the purchase

Well I tried to point that out to you that there wasn't such a code. But you simply didn't respond. As a result I just gave you all the codes that were there. to see whether something would stick.	Seeking more information, sending all the information available
I just tried every number until I got the right one but the process was kinda frustrating	Sending all the information available
Stopping the purchase and going to a different website	Quit the purchase proces
'well, then I am not buying it'	Not buying the product
I was trying all the numbers i saw on the screen	Sending all the information available
I thought I became stupid for not understandig	
Why do I not understand this problem and how can I make this chatbot more useful for me.	
I thought I'd better read everything first, and then started entering other numbers.	Sending all the information available
However, I did feel rushedChatting with the bot and trying again on a different site/device	Trying again
just giving you all the numbers	Sending all the information available
will I quit? maybe not because of all the wiring on my head.	Quitting the purchase process
therefore i just tried somethingyou were looking for an error number, which i could not find	
To ask you about the error number	Asking for help
Should I type in all the numbers or should I ask the bot more specifics about the number you wanted from me.	Asking for help vs. Sending all the information available
There were multiple numberswhich numbers should i read	Seeking help
Trying all different numbers one by one. Guessing which number would be the error number...	Sending all the information available
I got a bit fed up with it. But I wanted to resolve the problem so I just starting typing all the numbers that I saw to hopefully get the right number so that you could help me further.	Sending all the information available
Just typing out each and every possible numer	Sending all the information available
First of all asking you which number I needed specifically and secondly just trying every number.	Asking for help/more information

I wanted more information about which number i had to provide, so is pressed the 'i' first, but that was incorrect	Search for more information, clicking for more information
Quit the buying process and buy product somewhere else. And trying to figure out to solve the problem myself	Quitting the purchase process, postponing purchase
Not a specific action?	
Just send al the numbers i see	Sending all the information available
I clicked the i buttoni asked you where to find itand then i typed in everythingevery code i mean	Searching and asking for more information, sending all the information available
Just trying out every possible number I saw on screen and see which one you agree with, as you did not give further information when I asked which one you wanted.	Searching more information, sending all the information available
Asking you what to do	Asking for help/more information
nothing	
i wanted to ask someone personally instead of the robotand maybe i should of asked the robot for help but i did not think it was able to help me	Asking a human for help
As mentioned before, I considered asking the chatbot what it requires, after it could not find the receipt number. With regards to the experiment, it seemed a bit off that the purchase platform is mimicking the amazon interface, but does not behave t	Asking for more information
What do you mean by actions?I am afraid I am unable to provide you with further assistance. Please answer with the best of your abilities.Well I felt frustrated and irritated	
I was just thinking about what numbers to try. I was a little bit afraid that none of the numbers would work	