Chatbot says

"Sorry, I don't understand":

Recovery strategies for chatbot service failure

Claudette Greaves

A research report submitted to the Faculty of Management, University of the Witwatersrand, in partial fulfilment of the requirements for the degree of Master of Management in the field of Digital Business

Johannesburg, 2021

(Version 2.0)

ACKNOWLEDGEMENTS

Firstly, I would like to express my sincere gratitude to Prof. Gregory Lee, my supervisor, for affording me his time and accommodating me at late hours when required. I am most appreciative of his profound knowledge imparted and patience with my limited statistical skills. Patience is a well-respected trait of true educator. My learning curve has been steep under his supervision.

Secondly, I would like to thank my employer for funding my studies and also thank fellow colleagues who have assisted me in this journey. To Sanusha Naidoo for developing the chatbot scenarios to my specifications, and Conrad De Kock for your insights on the existing chatbot and UX design standards. To Anusha Mcneill, thank you for assisting me to role-play and record the call centre scenarios for participants to have a real-life call centre experience. To Alex Oloo for your thought leadership and support to make this a reality. To Helen Samodien and Ella Manana for assisting with arrangements for KLA to configure my survey with experimental scenarios for data collection. To Monique Lemmer for your support to overcome the compliance hurdles. To Gerda Naude and Clive Lees for assisting to source the survey target audience according to my requirements and ensure contact governance and data privacy controls were applied.

Thirdly, I would like to thank the 264 participants who completed the survey.

Most of all, I am greatly thankful to my family for their many sacrifices, understanding and support that made this milestone achievable. Your support and motivation in times of doubt and exhaustion has kept me determined to reach this achievement.

Giving thanks to the Almighty for keeping us safe during the most uncertain black swan event of our lifetime, the COVID 19 pandemic.

Claudette Greaves

DEDICATION

This study is dedicated to my late friend Bridget Camacho, who completed her Masters degree in Innovation while on chemotherapy. Her determination, strength and bravery has inspired me and will be remembered.

May this inspiration pass to my daughters, Dominique, and Gabrielle.

May you achieve anything you put your minds and hearts to.

ABSTRACT

AI Chatbots, especially those that are in early stages of development, can often be prone to 'failure' such as the inability to understand a customer question or to retrieve appropriate answers. This paper integrates the theories of expectation-confirmation, equity, and justice to construct an understanding behind customer's preferences for organisational recovery strategies when nascent chatbots fail in South African financial services. The study proposes that when a failure occurs with a chatbot, customer expectations, perceived inequity, and justice influence recovery of customer outcomes to pre-failure levels or even better. The research takes an experimental approach to examine the impact of different organisational recovery strategies on satisfaction, loyalty, and intent to re-use the chatbot in a banking case study.

KEYWORDS

chatbot; service failure; recovery strategies; justice; intent to re-use; loyalty

DECLARATION

DECLARATION

I, CLAUDETTE GREAVES, declare that this research report is my own work except as indicated in the references and acknowledgements. It is submitted in partial fulfilment of the requirements for the degree of Master of Management in the field of Digital Business at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination in this or any other university.

Name: CLAUDETTE GREAVES Signature:

Signed at EDENVALE

On the 27TH day of FEBRUARY 2022

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

AI	Artificial Intelligence
App	Mobile Banking Application
FAQ	Frequently Asked Questions
NLP	Natural Language Processing
ANOVA	Analysis of Variance
MANOVA	Multivariate Analysis of Variance
CFA	Confirmatory Factor Analysis

CHAPTER 1. INTRODUCTION

1.1 Purpose of the study

A South African financial service case study is performed to experiment and examine the impact of organisational recovery strategies on customer outcomes when nascent chatbot services fail.

1.2 Background of the study

Banking services have existed for centuries, traditionally conducted by human agents engaging with customers face to face via branches, or telephonically via call centres. The advent of technology has redefined the way customer engagement occurs, enabling more touchpoints for customers with greater flexibility, easier access to products and services, ease of use, and responsive to customer needs (Singh et al., 2019). Competition in financial services has made adaptation to the digital era an imperative for incumbent banks to survive. In the South African market, the recent emergence of digital only banks is evidence of digital disruption of banking services, creating further pressure for incumbent banks to adopt new technology advanced ways of work to remain customer-centric and reduce operating costs to survive (Camarate & C, 2018).

In this research, the use of chatbots for endpoint interaction with customers for banking services is explored. The 2018 Grand View Research report forecasts the chatbot market is to grow to $1.25 billion by 2025, with an average growth rate of 24.5% (Singh et al., 2019). Within financial services including banking and insurance, chatbots are expected to create value by improving customer satisfaction of interactions through digital touchpoints, as well as reducing operating costs associated with traditional human-assisted channels by increasing digital channel usage.

Financial chatbots are artificial intelligence (AI) conversational software programs that operate by users typing in a question in natural conversational language, and the software responds in a human-like simulated conversation. From a user perspective, financial chatbots enable smartphone users to perform banking services anywhere, anytime (Zumstein & Hundertmark, 2017). Mature financial chatbots can fulfil simple to complex tasks, from everyday cash management enquiries to advice for smarter financial decision making (Abe, 2016). Technological advancements in machine learning has created opportunities for chatbots to perform more complex tasks requiring intuition and richness (Poser et al., 2021).

Chatbots function by combining three crucial components. Firstly, chatbots require technology for message exchange with the customer. Secondly, chatbots require artificial intelligence which acts as the brain or knowledge base of the chatbot and enables understanding of natural language and decision making by decoding instructions. Lastly, chatbots require business processes designed to fulfil the core purpose of the chat to support banking servicing. This is done by accessing and sharing information in a secure manner that is compliant with privacy policies (Singh et al., 2019).

1.2.1 *Research problem*

The South African banking sector still has a nascent chatbot environment as both incumbent and digital banks have recently started experimenting with chatbots. In this early stage of development of banking chatbots in South Africa, chatbot knowledge bases are still very limited, resulting in limited conversational capabilities. These limitations in conversational capabilities may have a negative impact on user perception of chatbot technology, resulting in user frustration and discontinued use of the technology (Diederich et al., 2021). Conversation failure poses a risk of conversation loops in which the chatbot does not provide meaningful reply, leading to service failure and unsatisfactory customer which may negatively impact customer loyalty and retention (Ashktorab et al., 2019; Poser et al., 2021). Service failure and suboptimal customer experience creates a critical need for service recovery to restore customer satisfaction and complete fulfilment of the banking service.

Service recovery is the process of responding to a service failure to rectify customer dissatisfaction (Kelley & Davis, 1994). Parasuraman et al. (1991) argues that understanding customer expectations and responding suitably are pivotal in gaining a competitive advantage and suggests that recovery encounters which improve service can provide opportunities to swing customer frustration into loyalty. Research suggests that while good service recovery strategies can re-establish customer satisfaction, in contrast, poor recovery strategies can worsen customer dissatisfaction compared to immediately after the service failure (del Río-Lanza et al., 2009). Smith and Bolton (1998) suggests that up to twelve positive experiences are required to overcome one bad service experience, making it crucial for the most effective recovery strategies to be determined and implemented for a competitive advantage.

The suitability of service recovery techniques for the South African market is dependent on the digital divide in which experience with technology varies significantly amongst mobile banking application users.

Given the above the research problem becomes:

> **Understanding the dynamics of chatbot service failure in financial services and implementing effective recovery strategies.**

1.3 Research questions (RQ)

This research aims to explore the impact of chatbot failure on customer outcomes including customer satisfaction, loyalty, and intent to re-use the chatbot technology. The use of varied service recovery strategies is also explored to identify preferred service recovery types best suited for the South African market, which lead to improved customer outcomes. The dimensions of justice are also explored in this research as outlined in the literature review, due to the close correlation between justice dimensions and service recovery from extant literature.

In addition to the fundamental research questions, experience with technology is explored as a dependent variable, taking into consideration the wide digital divide in South Africa resulting in a broad range of tech savviness amongst South African banking customers which may affect the preferred service recovery type.

As service types requested via the chatbot may vary in importance to customers, salience of service type is explored as another dependent variable to determine whether tolerance of service failure in impacted, and whether preference of service recovery type is influenced.

This research therefore aims to answer the following research questions in a scenario-based study:

RQ1: Does chatbot failure without adequate recovery strategies lead to a decrease in customer outcomes?

RQ2: Do positive perceptions of recovery strategies improve customer outcomes over pre-failure levels?

- **RQ2a:** Do positive perceptions of recovery strategies improve brand loyalty?
- **RQ2b:** Do positive perceptions of recovery strategies improve intent to re-use the chatbot service for future banking services?
- **RQ2c:** Do positive perceptions of recovery strategies improve satisfaction?

RQ3: Which recovery strategy is most likely to meet customer expectations best leading to most improved customer outcomes over pre-failure levels?

- **RQ3a:** Does handover to a human agent lead to a higher increase in customer outcomes over pre-failure levels than other recovery strategies?
- **RQ3b:** Does a sense of distributive justice restoration or equity rebalancing partly explain the change in customer outcomes following successful recovery strategies?
- **RQ3c:** Does a perception of interactional justice interact with the relationship between the availability of a human agent handover and improved customer outcomes?
- **RQ3d**: Does a perception of procedural justice interacts with the relationship between the service recovery type and improved customer outcomes?
- **RQ3e:** Are more digitally native customers more tolerant of alternative non-human digital interfaces as an initial recovery strategy rather than going straight to human handover, than would non digital natives?
- **RQ3f:** Are customers more tolerant of service failure when asking the chatbot salient questions, than they would when asking the chatbot fundamental financial services questions?

1.4 Significance

Extant literature on human-supported services does not extend broadly to technology-supported services, as the nature of service failure and recoveries differs fundamentally. Recent studies have laid a foundation for service failure and recovery using chatbots, however application to the South African and emerging markets is very limited. This paper intends to contribute to this field of research in the South African context with unique digital divide. It is intended to provide valuable insights for the South African financial services industry to inform strategic direction and guide investment in terms of recovery strategies for financial chatbots. This study may also be extended to emerging markets with similar digital divide, especially in African countries with highly unbanked communities with low financial inclusion, positioning chatbots as a disruptive technology to transform banking business models (Abdulquadri et al., 2021).

1.5 Delimitations

i. Limited to text-based financial chatbots applied to banking service.
ii. Limited to active mobile banking app users, excluding other digital channels.
iii. Limited to justice needs and excludes other basic needs which contribute to brand loyalty, such as security and self-esteem needs.
iv. Scenario-based approach used in research design cannot fully replicate all activities involved in real-life experience triggered by service failure.

1.6 Definition of terms

The paper refers to the following terms:

- *Chatbot* is an artificial intelligence based conversational agent capable of mimicking a conversation with a human using natural language processing (NLP) (Schuetzler et al., 2018).
- *Service failure* is a problem or error experienced during a service encounter with an organisation (Maxham, 2001).
- *Service recovery* is the process or action of responding to a service failure in order to rectify customer dissatisfaction (Kelley & Davis, 1994).

1.7 Assumptions

The following assumptions are applicable to this study:

- The handover efficiency from chatbot to human call centre agent and human online agent is adequate and represents a real-life situation.
- It is assumed that controlled human contact scenarios such as call centre recorded responses simulate a real-life experience.

1.8 Chapter Outline

This study was inspired by the digital age in which technology enabled chatbots have created opportunities to transform the financial service arena. The journey to reach chatbot maturity is accompanied by service failure and dissatisfaction, creating a need for service recovery to keep customers satisfied, loyal, and frequently using chatbots for banking services.

CHAPTER 2. LITERATURE REVIEW

2.1 Introduction

The literature reviewed encompasses chatbot service failure and recovery studies based on the underlying theories of expectation-confirmation, equity and justice theory, and the impact on customer outcomes including satisfaction, loyalty, and intent to re-use the service.

2.2 Chatbot service failure and recovery

2.2.1 *Chatbot service failure*

While chatbot technology has made headway since their introduction, many examples of high profile chatbot failures have been documented. A recent example includes Microsoft's AI Chatbot, Tay, that was decommissioned after posting offensive tweets relating to feminism and drug consumption (Review, 2016). Another more recent example includes Facebook chatbot shut down in 2017 after conversing in its own language not understood by humans (Tech2, 2022).

Literature suggests that service requests exceeding a chatbots limited language understanding often lead to conversational loops or breakdowns, leaving the customer's service request unresolved (Corea et al., 2020). Recent studies on complex chatbot conversational interactions suggest a range of repair or recovery strategies from digital self-repair methods to hybrid service recovery with handovers to human call centre or virtual agents (Ashktorab et al., 2019; Magnusson & Rånnerud, 2019; Poser et al., 2021). In the context of banking services, chatbot failure without adequate recovery can severely impact customer satisfaction, loyalty, and intent to re-use the chatbot technology frequently in future.

2.2.2 *Service recovery*

Boshoff (1999) states that after a customer has been disappointed by service failure, an effective service recovery strategy is vital to restore customer satisfaction. Further literature suggests that service recovery is a critical strategy for service providers to achieve customer retention and loyalty. Customer perceptions of service recovery is influenced by what is done to recovery, as well as how it is done (Andreassen, 2000; Tax et al., 1998; Wirtz & Mattila, 2004).

2.3 Customer satisfaction and loyalty

Satisfaction is derived from the psychological state in purchasing behaviour that is emotional, favourable, and subjective (Jung et al., 2017). Satisfaction has been studied extensively in marketing as a key construct to determine subsequent customer behaviour. Studies suggest the more satisfied customers are, the more loyal they are likely to be. Leading to repurchase behaviour. Dissatisfied customers, on the other hand are more likely to discontinue the use of a product or service or substitute it (Dwivedi et al., 2012; Oliver, 1999). Customer acquiring costs may be as high as five times the cost of retaining existing customers, supporting the underlying rationale for companies placing a primary focus on keeping customers satisfied to retain them (Dwivedi et al., 2012).

In a competitive markets, service providers who strive for customer loyalty need to make every effort to reach beyond satisfying customers to delighting them (Schneider & Bowen, 1999). Customer delight is described as the experience of pleasant surprise which results from unexpected value (Severt, 2002). Oliver (1999) defines loyalty as a pledge to re-patronise a preferred providers service or product, irrespective of competitor marketing efforts enticing defection.

Satisfaction of the recovery process can directly influence customer behavioural intentions such as loyalty, and as a result, the intent to repurchase or re-use a service. The *service recovery paradox* suggests that customers who have experienced a failure accompanied with appropriate recovery action exhibit higher customer satisfaction outcomes when compared to pre-failure customer outcomes (Smith & Bolton, 1998).

2.4 Underlying theories

Three theoretical frameworks have been referenced extensively in service recovery literature, namely justice, equity, and expectation-confirmation theory.

2.4.1 *Expectation-confirmation theory*

Expectation-confirmation or expectation disconfirmation theory (Oliver, 1980) has been extensively used as a foundation for extensive marketing and consumer behaviour research to explain customer repurchase intentions. Expectation – confirmation theory posits that an individual forms a perception of a service or product prior to purchase, and the level of customer satisfaction is determined by their assessment of preceding expectations of the service or product against its performance. This leads to an outcome that is either better than expected (a positive disconfirmation) or worse than expected (a negative disconfirmation).

Based on this theory, a consumer who has experienced a service failure will have an expectation of recovery, and positive or negative disconfirmation of recovery efforts will influence the consumers attitude towards the service or product, and in turn result in behaviours such as loyalty and intent to repurchase or re-use the service (Lai & Chou, 2015).

A negative disconfirmation may result in loss of customers or negative word of mouth, while a positive disconfirmation may result in an elevated perception of service quality, loyalty or continued patronage in terms of repurchase or re-use of a service (Dwivedi et al., 2012).

2.4.2 *Equity theory*

Adams (1963) equity theory explains human motivation in exchange relationships as individuals seeking a fair balance of inputs and outputs relative to others (a referent). Feelings of equity result when there is a perceived balance of inputs and outputs compared to the referent. In contrast, feelings of inequity result when inputs and outputs differ compared to the referent, and may motivate the individual to act or modify this imbalance (Rasooli et al., 2019; Ryan, 2016).

Equity theory has been applied frequently in retailing and consumer services mostly in articles on service failure, justice perception, price examination and service failure compensation (Buse & Hassan, 2019). In the service arena equity theory refers to a consumers psychological comparison of inputs (participant contribution to the exchange e.g., time, effort, money) and outputs (tangible or intangible results e.g., fulfilled service request), comparing the participants input and output ratio with that of a referent (e.g., another customer) (Lii et al., 2018).

Equity theory posits that customers will be satisfied when the individual's ratio of outcomes to inputs is greater than that of the referent. In contract, when this ratio is lower than that of the referent, this perceived inequity results in frustration or anger against the service provider which may cause customer dissatisfaction (Buse & Hassan, 2019).

Applying equity theory to chatbot conversation failure may create feelings of inequity, requiring good service recovery strategies that may turn a negative outcome into a positive one and restore satisfaction.

2.4.3 *Justice theory*

Justice theory has origins in Adams (1965) equity theory and centres on the basic human need to be treated fairly (i.e., justice). Research in social psychology and philosophy suggests that justice is central to relationships and refers to an implicit "psychological contract" between people as individuals, organisations, or society. Justice theory has been extensively applied as an underlying framework for service recovery given its central nature of relationships and social exchange in the services field (Cai & Qu, 2018; Lii et al., 2018; Petzer et al., 2017). Service failure can be a violation of justice, while recovery action can restore justice.

Justice theory evaluates fairness by three key dimensions, namely distributive, procedural, and interactional justice (Smith & Bolton, 1998; Tax & Brown, 1998).

In a service recovery context,

Distributive justice evaluates a customers perceived fairness of the outcome of the exchange (e.g. service encounter) (Jasso et al., 2016; Severt, 2002).

Interactional justice evaluates the perceived fairness of the interpersonal treatment received in the engagement, generally based on human behavioural attributes such as empathy, politeness, honesty and explanation (Severt, 2002; Yani-de-Soriano et al., 2019).

Procedural justice evaluates the process or actions by which the outcome of exchange is reached, making judgement based on attributes such as process speed, timing and flexibility and responsibility (Tax & Brown, 1998).

2.4.4 *Theoretical base*

The three theoretical frameworks of expectation-confirmation, equity and justice form a sound theoretical foundation for this research as equity and justice considerations are antecedent to customer satisfaction, loyalty and repatriation behaviour such as repurchase or re-use of a service (Oliver & Swan, 1989).

Building on expectation-confirmation theory, equity theory and justice theory this paper considers the impact of organisational recovery strategies on customer outcomes including, satisfaction post recovery, customer loyalty, and intent to re-use the chatbot service.

2.5 Research question 1

Applying equity theory, chatbot conversation failure can be perceived as inequity. Consequently, the following hypothesis is proposed:

2.5.1 *Hypothesis 1*

Inequity is a mediating factor to the decline in customer outcomes following chatbot failure.

2.6 Research question 2

2.6.1 *Hypothesis 2*

Service recovery theory suggests that positive perceptions of recovery strategies improve customer outcomes over pre-failure levels. A scenario-based experimental study aims to test the following hypotheses proposed for each customer outcome tested:

H2a: Positive perceptions of recovery strategies improve brand loyalty.

H2b: Positive perceptions of recovery strategies improve intent to re-use the chatbot service for future banking services.

H2c: Positive perceptions of recovery strategies improve satisfaction.

2.7 Research question 3

2.7.1 *Hypothesis 3*

Service recovery research suggests that suitable service recovery efforts can pivot customer frustration resulting from service failure into loyalty. Research also suggests that handovers from chatbots to human service agents are preferred (Poser et al., 2021). The scenario-based experiment aims to test the following proposed hypothesis:

H3a: Handover to a human agent leads to a higher increase in customer outcomes over pre-failure levels than other recovery strategies.

Justice theory has been foundational in human-centred service recovery research, however limited research is available on technology-based service failure and recovery. Consequently, the following hypotheses are proposed to determine the interaction of chatbot service recovery with the dimensions of justice.

H3b: A sense of distributive justice restoration or equity rebalancing partly explain the change in customer outcomes following successful recovery strategies.

H3c: Perceived interactional justice interacts with the relationship between the availability of a human handover and improved customer outcomes.

H3d: Perceived procedural justice interacts with the relationship between the service recovery type and improved customer outcomes.

Experience with technology ranges broadly across the South African market due to the digital divide, and may influence customer preference for human-assisted versus non-human digital-based recovery strategies The following hypothesis is proposed:

H3e: Digital natives are more tolerant of alternative non-human digital interfaces as an initial recovery strategy rather than going straight to call centre, than would non digital natives.

Tolerance of service failure may be influenced by the level of perceived importance of service request uttered for chatbot assistance. The experimental study includes scenarios with varied level of salience to test the following proposed hypothesis:

H3f: Customers are more tolerant of service failure when asking the chatbot salient questions, than they would when asking the chatbot fundamental financial services questions.

2.8 Conclusion of Literature Review

Appendix E includes a consistency table summarising the ten hypotheses for this study.

2.9 Analytical framework

This framework integrates the underlying theories of justice, expectation confirmation and equity, to illustrate their impact of chatbot service failure and the implementation of service recovery strategies to restore customer outcomes.

2.9.1 *Theoretical Framework*

This model conceptualises the path from a customer's pre-conceived expectations of a service, experiencing chatbot failure, and reaching customer outcomes through recovery action and restoration of equity and justice dimensions.

Based on the underlying literature, service failure assessed against a customers preconceived expectations will result in feelings of inequity or unfairness, leading to customer frustration which in turn result in negatively influence customer satisfaction, branch loyalty and intent to re-use the chatbot technology. The implementation of suitable recovery strategies restores equity or distributive justice, procedural and interactional justice, which in turn positively influence customer satisfaction, loyalty, and intent to re-use the chatbot technology for future servicing.

2.9.1 *Conceptual Framework*

The conceptual framework models the path from preconceived customer expectations and service failure to loyalty and intent to re-use the chatbot technology by restoration of equity and justice through service recovery.

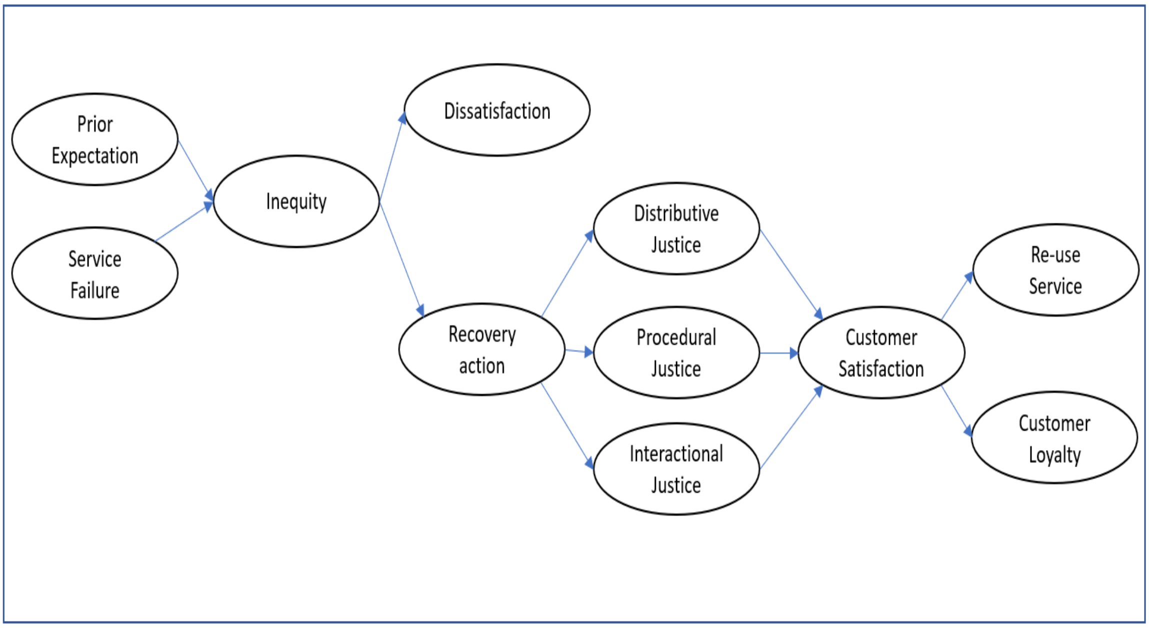


Figure 1: Conceptual model of the recovery path from service failure to customer loyalty and re-use of service

CHAPTER 3. RESEARCH METHODOLOGY

The research methodology selected was centred on a real-life scenario-based experiment to understand the dynamics of chatbot service failure to enable implementation of preferred recovery strategies to achieve the desired customer outcomes. A case study was selected using a South African incumbent bank using chatbot technology in the mobile banking app.

3.1 Research approach

This research used an incumbent bank as a case study for a scenario-based experiment simulating real-life experience of chatbot failure, followed by different service recovery strategies. The experiment was limited to a chatbot in the mobile banking app primarily used by customers for service queries. Chatbot service failure in this context is defined as the chatbot failing to provide a quality response to a service question utterance, often leading to a conversation loop with no positive outcome. With an unsuccessful answer, the customer is unable to perform the necessary service on the mobile app, and despite frustration experienced, most like used an alternative human-assisted channel, increasing the cost to service and limiting the app from becoming a preferred channel for servicing.

The experiment was limited to three controlled service queries of varying importance to customers, and included the following three questions uttered by a customer:

1. I would like to increase my debit card daily limit
2. I would like to delete a beneficiary on my app
3. I would like to see my vehicle finance account in my app

For each service query listed above, 6 different controlled scenarios were developed. This includes once scenario in which chatbot failure had no recovery, as well as five scenarios where chatbot failure was immediately followed by a different recovery strategy. Table 1 lists a description of the six scenarios developed per service query type listed above.

#	Recovery Strategy	Description
Failure	Failure – No Recovery	Service failure with no recovery in the chatbot simulation. The service ends with no recovery to aid on fulfilling the requested service via the chatbot.
1	Digital Guided – with Procedures	A fully digital journey is simulated with selectors leading to a self help procedure guide provided for the user to self serve. The procedure is in the form of a step by step guide displayed in mobile App. This is described simply in the scenarios as "tell me how".
2	Defer to Call Centre	Automated referral to a human call centre agent. A recorded conversation with a human call centre agent was used in this scenario to simulate a real-life call centre conversation.
3	Defer to Online Agent	Automated referral to a human online agent. The scenario simulates an online agent icon with timed conversation responses to simulate a real life two way virtual conversation via chatbot in mobile App.
4	Digital Navigation	A fully digital recovery with selectors, leading to automated navigation to the applicable service screen to aid fulfilment of the required service. This is described simply in the scenarios as "take me there".
5	Digital Guided – with Recovery options	A digital recovery method providing selectors to the user to choose their preferred recovery path

Table 1: Recovery strategy descriptors

Each service scenario was built using a standard design template according to Bank X brand standards to control other variables.

Based on the literature review, measuring customer outcomes in a service encounter is most optimal in real life situations to capture customer feelings as they are experienced and influence customer behaviour (Smith & Bolton, 1998). Due to the complex nature of measuring real-life customer outcomes triggered by service failure, a scenario-based experimental design approach was selected based on appropriateness.

This experimental service experience designed was coupled with pre- and post-experience surveys to measure customer preferences for recovery strategies for the nascent chatbot service failure.

Chatbot scenarios were developed for service failure as well as five recovery strategies including namely 1) digital guided – with procedures; 2) defer to contact centre agent; 3) defer to online agent; and 4) digital navigation, and 5) digital guided – with recovery options. Service requests asked in the chatbot were controlled and limited to 3 types of varying salience.

The research approach used for this research consisted of the three stages depicted in Figure 2 below:

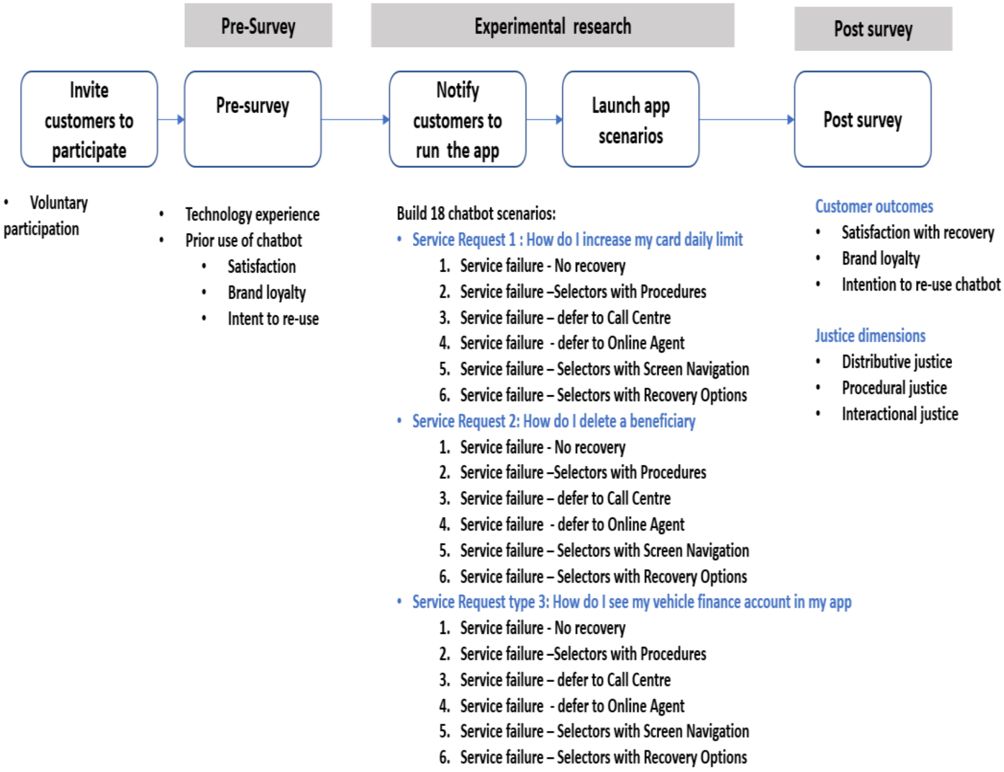


Figure 2: Scenario-based research approach to understand customer's preference for recovery strategies

3.2 Research design

A three-stage process was used that was centred on an experimental chatbot in mobile app using 18 experimental scenarios developed with varied 1) salience of service task request to the chatbot; 2) failed or successful outcome; and 3) recovery strategy type. The experimental mobile app scenarios tested five recovery strategies including 1) digital guided – with procedures; 2) defer to contact centre agent; 3) defer to online agent; and 4) digital navigation, and 5) digital guided – with recovery options.

The research measured the following eight constructs, namely 1) experience with chatbots; 2) technology experience; 3) distributive justice; 4) interactional justice; 5) procedural justice; 6) brand loyalty; 7) satisfaction; and 8) intent to re-use the chatbot. Satisfaction, loyalty, and intent to re-use were measured in both pre-and post-survey to measure a change in customer outcomes subject to the scenario experience.

3.3 Data collection methods

As discussed in the research design, data collection consisted of three stages conducted in sequence.

Stage one was pre-data collection. The data collection method was a survey that measured four variables, namely 1) technology experience, 2) prior use of the nascent chatbot; 3) intent to re-use the chatbot; 4) brand loyalty and 5) satisfaction with the chatbot.

Stage two exposed participating customers to one of the eighteen mobile app scenarios developed. Each participant experienced one service simulation via their mobile phone. Research scenarios were randomly allocated to participants and recovery strategy preference data was collected. Participants were randomly selected from a stratified sample of Bank X's customer base of active mobile app users across demographics.

Stage three collected data via a survey post experience the scenario. Six constructs measured include 1) intent to re-use the chatbot; 2) satisfaction; 3) loyalty; 4) distributive justice; 5) interactional justice; and 6) procedural justice.

The three stages consisting of pre-survey, experimental scenarios, and post survey were aggregated and configured into one survey by KLA, a market research firm that performed quantitative surveys with customers on behalf of Bank X. An introductory letter was sent to the target audience prior to sending the survey, which included terms relating to data privacy. Data was collected by KLA from 10th January to 18th February 2022, with a total of 264 complete responses recorded. The disbursement of non-material competition value for respondents was also administered and managed by KLA. Introductory letter template is included in Appendix D.

3.4 Population and sample

3.4.1 *Population*

Two categoric population groups were included in this study:

Firstly, this research was applicable to customers of all South African retail incumbent banks with a nascent chatbot, having a similar customer base.

Secondly, the study can be generalised to other similar emerging market retail banks with native chatbots, or banks considering using chatbots in future to reach communities unable to access physical banking channels.

Further extrapolation to global mature markets is not recommended despite the advanced nature of banking in South Africa, as the digital divide may affect recovery preferences.

3.4.1 *Case site and target population*

Given that this is a case study, the South African banking organisation used as the case site was referred to as Bank X. The total case site was the sample frame, including all Bank X retail customers who could actively use the mobile banking app.

3.4.2 *Sample and sampling method*

The sampling frame consisted of 22 000 active mobile banking customers who were approached to participate in this study. The sample was selected to represent Bank X's customer base demographics and geographies, with random allocation of chatbot scenarios to each participant. The demographic profile of volunteering respondents was dependent on who answered and who downloaded the app, however by virtue of the sampling nature it is hoped that it mirrors the Bank X retail population.

The sampling method involved sending an invitation to voluntarily participate in the experiment. Given the nature of the experiment, there was no feasible method to predict the response rate of participants. 264 complete responses were collected, indicating a response rate of 1.2% complete responses after data clean-up withing the data collection period.

3.5 The research instruments

3.5.1 *Pre-survey*

As discussed under research design stage one was measured via a pre-survey using the four constructs, namely 1) technology experience; 2) prior use of Bank X's chatbot; 3) intent to re-use the chatbot; 4) brand loyalty and 5) satisfaction. Brand loyalty and satisfaction were only measured for participants who have used Bank X's chatbot previously. Appendix A lists the full set of scale items.

Technology experience: This was measured using 3 self-reported scale items, including 2 items from Sudzina (2015) and 1 from Ashktorab et al. (2019), applying a 5-point Likert scale from strongly agree to strongly disagree. One example of a scale item used is "*I see myself as an advanced technology user*".

Prior use of Bank X's chatbot: This was measured with 1 scale item "I have used the Bank X chatbot before and have attempted to ask the chatbot a question". This item was measured using a 3-point scale including agree; not sure and disagree.

Intent to re-use the chatbot and satisfaction scale items were only measured for participants who have indicated prior use of Bank X's chatbot.

Intent to re-use: Self-reported intent to re-use the chatbot was measured using 2 scale items adapted from Ilias et al. (2014) and used a 5-point Likert scale ranging from strongly agree to strongly disagree. One example of a scale item is "Based on my experience, I would use Bank X's chatbot frequently in future".

Customer satisfaction: This was measured using 3 scale items from del Río-Lanza et al. (2009) and used a 7-point Likert scale from totally disagree to totally agree. One example of a scale item is “I am happy with the way my service request was handled”.

Brand loyalty: Loyalty was measured using a 7 item scale items from Bobâlcă et al. (2012) including three types of loyalty measures, namely affective, conative and action loyalty and used a 7-point Likert scale from totally disagree to totally agree. Affective loyalty was assessed using 2 scale items, including sample item “I am pleased to use Bank X instead of other financial services brands”. Conative loyalty was assessed using 2 scale items, including sample item “I intend to use other products from Bank X”. Action loyalty was assessed using 3 scale, items including sample item “I recommend Bank X to those who ask my advice”. Brand loyalty was measured for all participants, irrespective of prior exposure to Bank X’s chatbot.

All questions in the pre-survey had compulsory responses, with exception of respondents with prior experience of Bank X’s chatbot who were asked additional questions on intent to re-use and satisfaction based on prior use of the chatbot. Based on this control question, 162 respondents had no prior exposure to Bank X’s chatbot, with missing values for intent to re-use and satisfaction. 102 respondents had prior experience of Bank X’s chatbot, resulting in no missing data for all scale items in both pre- and post-survey.

3.5.2 *Experiment instrument*

Customers were invited to engage in a mobile app scenario in which they imagine using the chatbot in an experimental mobile banking app. Participating customers were presented with one of eighteen randomly selected chatbot scenarios to experience. All scenarios include a pre-typed controlled service request simulating the customer asking the chatbot for assistance. In each scenario the chatbot service failed, followed by no recovery, or one of 5 varied recovery strategies which led to a successful service outcome after following the prompts in the controlled scenario.

The user experience and interface of all scenarios designed followed the standard design principles and existing design of Bank X mobile app, keeping this generic design consistent across all scenarios to control variation.

Table 2 below lists the eighteen experimental app scenarios included in the experiment.

Scenario #	Service request #	Chatbot service request question	Recovery strategy #	Recovery strategy
0	1	How do I increase my debit card daily limit?	Service Failure	Service failure - No Recovery
1	1	How do I increase my debit card daily limit?	Recovery 1	Digital Guided– with Procedures
2	1	How do I increase my debit card daily limit?	Recovery 2	Defer to Call Centre
3	1	How do I increase my debit card daily limit?	Recovery 3	Defer to Online Agent
4	1	How do I increase my debit card daily limit?	Recovery 4	Digital Navigation
5	1	How do I increase my debit card daily limit?	Recovery 5	Digital Guided – with recovery option selectors
6	2	How do I delete a beneficiary	Service Failure	Service failure - No Recovery
7	2	How do I delete a beneficiary	Recovery 1	Digital Guided– with Procedures
8	2	How do I delete a beneficiary	Recovery 2	Defer to Call Centre
9	2	How do I delete a beneficiary	Recovery 3	Defer to Online Agent
10	2	How do I delete a beneficiary	Recovery 4	Digital Navigation
11	2	How do I delete a beneficiary	Recovery 5	Digital Guided – with recovery option selectors
12	3	How do I see my vehicle finance account in my app?	Service Failure	Service failure - No Recovery
13	3	How do I see my vehicle finance account in my app?	Recovery 1	Digital Guided– with Procedures
14	3	How do I see my vehicle finance account in my app?	Recovery 2	Defer to Call Centre
15	3	How do I see my vehicle finance account in my app?	Recovery 3	Defer to Online Agent
16	3	How do I see my vehicle finance account in my app?	Recovery 4	Digital Navigation
17	3	How do I see my vehicle finance account in my app?	Recovery 5	Digital Guided – with recovery option selectors

Table 2: Experimental app scenarios included in research design

Figure 3 below depicts an example of the experimental app chatbot scenario used for data collection.

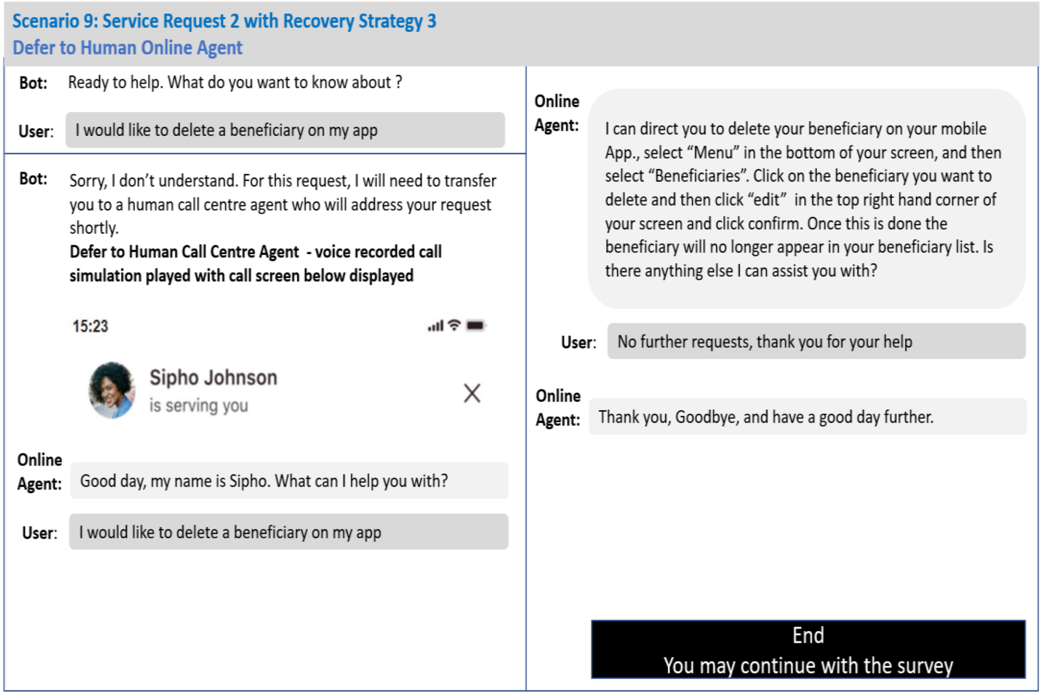


Figure 3: Example of an experimental app chatbot scenario illustrating a service failure with a recovery strategy

Appendix B depicts a visual illustration of the eighteen-research design experimental chatbot scenarios labelled 0 to 17.

A distribution of responses across experimental scenarios is presented in Table 3 below.

Chatbot Scenario	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Request 1 Defer to Call Centre	14	5,3	14	5,3
Request 1 No Recovery	16	6,06	30	11,36
Request 1 Defer to Online Agent	15	5,68	45	17,05
Request 1 Digital Guided - Options	10	3,79	55	20,83
Request 1 Digital Guided - Procedures	16	6,06	71	26,89
Request 1 Digital Navigation	15	5,68	86	32,58
Request 2 Defer to Call Centre	17	6,44	103	39,02
Request 2 No Recovery	18	6,82	121	45,83
Request 2 Defer to Online Agent	17	6,44	138	52,27
Request 2 Digital Guided - Options	15	5,68	153	57,95
Request 2 Digital Guided - Procedures	18	6,82	171	64,77
Request 2 Digital Navigation	12	4,55	183	69,32
Request 3 Defer to Call Centre	9	3,41	192	72,73
Request 3 No Recovery	12	4,55	204	77,27
Request 3 Defer to Online Agent	15	5,68	219	82,95
Request 3 Digital Guided - Options	14	5,3	233	88,26
Request 3 Digital Guided - Procedures	21	7,95	254	96,21
Request 3 Digital Navigation	10	3,79	264	100

Table 3: Distribution of respondents

3.5.3 *Post-survey*

Post exposure to the experimental app, stage three measured the following six constructs via survey, namely 1) intent to re-use; 2) customer satisfaction; 3) brand loyalty; 4) distributive justice; and 5) interactional justice; 6) procedural justice; and 7) service request salience. Appendix C lists the full set of scale items.

Intent to re-use: Scale items used in the pre-survey were repeated to perform comparative testing pre and post the service encounter. One example of a scale item is “Based on my experience, I would use the Bank X chatbot frequently in future”.

Customer Satisfaction: Scale items used in the pre-survey were repeated to enable a comparison of satisfaction before and after the simulated service encounter. One example of a scale item is "*My service request was resolved to satisfaction*".

Brand loyalty: Scale items used in the pre-survey were repeated to compare loyalty pre and post experiencing the simulated service encounter. One example of a scale item is "I consider Bank X my first choice when buying financial products and services".

Distributive justice: This was measured using 2 scale items adapted from Severt (2002). All justice dimensions were measured with a 7-point Likert scale ranging from totally agree to totally disagree. One sample item is "*I got the outcome I expected*".

Interactional justice was measured using 3 scale items adapted from del Río-Lanza et al. (2009) and Maxham and Netemeyer (2003). One example of a scale item is *"Bank X showed a real interest in my service request".*

Procedural justice was measured using 5 scale items adapted from del Río-Lanza et al. (2009). One example of a scale item is "*Bank X tried to solve my service request as quickly as possible*".

Service request salience: This scale item measured the self-reported level of importance of the 3 selected service tasks. A 5-point Likert scale was used ranging from not important to very important. One example of a scale item is "Rate the level of importance of the service requests by importance to you. Increase my card point of sale limit in my mobile banking app".

3.6 Procedure for data collection

Figure 2 above illustrates a flow diagram of the experimental research design and data collection procedure. Data is collected via a pre-survey prior to exposing participants to the chatbot scenario. Post experiencing the chatbot experiment, data is collected via a post-survey.

3.7 Data analysis and interpretation

Data collected was analysed using SAS OnDemand (SAS ODA) for academics.

Firstly, internal consistency and reliability of multi-item factor structures was tested using Cronbach Alpha (Tavakol & Dennick, 2011).

Secondly, structural equation modelling using a sequence of Confirmatory Factor Analysis (CFA) was used in SAS ODA to examine for issues such as convergent and divergent validity.

Lastly, causal effect modelling was performed using individual samples t-test and Analysis of Variance (ANOVA) to measure the relationships where the outcome measures were dependent variables.

3.8 Validity and reliability

3.8.1 *External validity*

The results were likely generalisable across South African retail banks, as the retail banking space is quite standardised. In addition, it may be generalised across similar emerging market retail banks with similar digital divide (Eren, 2021; Mogaji et al., 2020).

3.8.2 *Internal validity*

The experimental nature of the methodology assisted to control many factors that may have resulted in poor validity in service failure and recovery measurement, and therefore produced consistent, reliable, and trustworthy results (Wirtz & Mattila, 2004). The validity was also boosted by the strength of the initial construct validity testing using CFA.

3.8.3 *Reliability*

The reliability of the scale items was tested through Cronbach Alpha, an internal reliability statistical measurement. The experimental app and the app response mechanisms were highly standardised which facilitated reliability. This included control of the service request question posed in the free text conversation box. Controlled response mechanisms also limited the flow of each scenario to simulate to intended outcome.

The Cronbach Alpha values for each dimension were high, with α >.78. Based on this test, all variables were used for further analysis. Table 4 presents a list of Cronbach Alpha values for pre-and post-experiment survey measures.

	Cronbach Alpha	N
Pre-survey		
Tech experience	0.86	264
Intent to re-use 1	0.78	102
Satisfaction 1	0.97	102
Loyalty 1	0.96	264
Post-survey		
Intent to re-use 2	0.82	264
Satisfaction 2	0.95	264
Loyalty 2	0.97	264
Distributive justice	0.94	264
Interactional justice	0.96	264
Procedural justice	0.96	264

Table 4: Cronbach's Alpha coefficients of pre- and post-survey latent variables

3.9 Confirmatory factor analysis

A sequence of CFA was performed using SAS ODA. A first CFA was performed on pre-survey measures technical experience and loyalty 1. A second CFA was performed for pre-survey measures with missing values, including satisfaction1 and reuse 1. A third CFA was then performed for all post-survey measures including reuse 2; satisfaction2, loyalty2; distributional justice; interactional justice and procedural justice. Refer to Table 5 for fit statistics of each CFA performed.

3.9.1 *Confirmatory Factor Analysis Fit statistics*

		Tech Experience Loyalty1	***Satisfaction 1 Reuse 1***	***Satisfaction 2 Loyalty 2 Reuse 2 Distributive Justice Interactional Justice Procedural Justice***
Absolute Index	*Chi-Square*	102.7361	10.0896	487.1527
	Chi-Square DF	34	4	194
	Pr > Chi-Square	<.0001	0.0389	<.0001
	Standardized RMR (SRMR)	0.0324	0.0143	0.0223
Parsimony Index	*RMSEA Estimate*	0.0877	0.1228	0.0758
	RMSEA Lower 90% Confidence Limit	0.0685	0.0253	0.0674
	RMSEA Upper 90% Confidence Limit	0.1074	0.2198	0.0842
Incremental Index	*Bentler Comparative Fit Index*	0.9731	0.9906	0.97
	Bentler-Bonett Non-normed Index	0.9643	0.9764	0.9643

Table 5: Confirmatory Factor Analysis Fit Summary

Table 5 documents good fit statistics for tech experience and loyalty for the pre-survey, and good fit for all post survey variables. (Ch-square p <0.0001; SRMSR < 0.05; RMSEA <.88; Bentler Comparative Fit Index > 0.95 and Bentler-Bonett Non-normed Index > 0.95). Pre-survey satisfaction and intent to reuse displayed moderate fit (Ch- square p <0.05; SRMSR < 0.05; RMSEA >.10; Bentler Comparative Fit Index > 0.95 and Bentler-Bonett Non-normed Index > 0.95).

3.9.1 *CFA standardised path summary*

Table 6 presents a summary of path coefficients for CFA's performed

Relationship	*Estimate*	*t Value*	*p*
Tech_Experience	0,87	37,14	<.0001
Tech_Experience	0,90	41,01	<.0001
Tech_Experience	0,77	26,04	<.0001
Loyalty1	0,90	70,22	<.0001
Loyalty1	0,92	83,24	<.0001
Loyalty1	0,91	76,11	<.0001
Loyalty1	0,84	42,10	<.0001
Loyalty1	0,92	80,35	<.0001
Loyalty1	0,86	49,27	<.0001
Loyalty1	0,88	56,39	<.0001
Reuse1	0,89	29,39	<.0001
Reuse1	0,95	37,98	<.0001
Satisfaction1	0,94	68,74	<.0001
Satisfaction1	0,98	140,80	<.0001
Satisfaction1	0,97	108,30	<.0001
Reuse2	0,93	63,60	<.0001
Reuse2	0,91	58,36	<.0001
Satisfaction2	0,95	140,70	<.0001
Satisfaction2	0,96	165,30	<.0001
Satisfaction2	0,96	148,00	<.0001
Loyalty2	0,91	80,01	<.0001
Loyalty2	0,94	115,60	<.0001
Loyalty2	0,95	134,10	<.0001
Loyalty2	0,92	92,04	<.0001
Loyalty2	0,95	151,40	<.0001
Loyalty2	0,93	99,38	<.0001
Loyalty2	0,94	126,10	<.0001
DistJustice	0,98	204,70	<.0001
DistJustice	0,98	207,70	<.0001
InterJustice	0,93	98,47	<.0001
InterJustice	0,95	143,00	<.0001
InterJustice	0,97	183,60	<.0001
ProcJustice	0,93	96,82	<.0001
ProcJustice	0,92	88,50	<.0001
ProcJustice	0,91	83,65	<.0001
ProcJustice	0,95	149,20	<.0001
ProcJustice	0,96	159,30	<.0001

Table 6: Measurement Model: Confirmatory Factor Analysis Path Coefficients

Confirmatory factor analysis of a 37-item, 10 factor structure measurement model showed that all scale items loaded with good fit on the relevant latent variables

(P<0.0001 for all variables). This confirms that the measurement model has satisfactory fit for causal effect analysis for hypothesis testing.

3.10 Limitations

The use of scenarios does not fully replicate all activities involved in real-life experience triggered by service failure. Video recordings were used to simulate a real-life call centre conversation. Timed pre-scripted text conversations were used to simulate an online agent conversation.

3.11 Ethical considerations

This study is classified as one of minimal risk, as the only conceivable risk is one of inconvenience. Customer participation in the chatbot experiment and surveys followed the Bank X standard process for customer research. Participation was voluntary, and withdrawal from the study was available to be done freely at any time, at no disadvantage. All participant data was managed according to Bank X's data management policy. A non-material competition value was offered as incentive to complete the survey for volunteering participants.

CHAPTER 4. RESULTS

4.1 Introduction

From the confirmatory factor analysis, factor scores were derived for the latent variables for the 264 respondents in the study sample. These scores were used in causal effect analysis instead of latent variable means. Causal effect analysis was performed using one-way univariate analysis of variance (ANOVA) and individual samples t-test for hypothesis testing.

One way ANOVA and individual samples t-test was selected as the appropriate methods of analysis to determine the difference between sample means for different scenarios or groups. ANOVA is commonly used in market research in the service arena to test hypothesis to determine how different groups of consumers respond to surveys, or how the same group responds to different experimental scenarios (Magnusson & Rånnerud, 2019; Mattila, 1999). Similarly, a t-test was used to determine if satisfaction, loyalty and intent to re-use significantly varies pre and post experience of the experimental service encounter scenario.

In this experimental research, the null hypothesis (H_0) states that the means of customer outcomes before and after experiencing the different experimental service encounter scenarios are equal, implying no difference observed. The alternative hypothesis (H_A) states that the means of customer outcomes before and after experiencing the different experimental service encounter scenarios are significantly different.

4.2 Correlation Analysis

Table 7 includes the Pearson correlation coefficients for latent variables using factor scores

	SD	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
Tech_Experience	1.00	1.00									
Loyalty1	1.07	.28***	1.0								
Reuse1	1.00	.29**	.5***	1.0							
Satisfaction1	1.01	.25**	.54***	.83***	1.00						
Reuse2	1.04	.32***	.49***	.57***	.54***	1.00					
Satisfaction2	1.07	.26***	.45***	.44***	.48***	.85***	1.00				
Loyalty2	1.09	.22***	.76***	.36***	.40***	.67***	.69***	1.00			
Distributive Justice	1.11	.18**	.49***	.39***	.41***	.79***	.84***	.77***	1.00		
Interactional Justice	1.10	.22***	.60***	.40***	.41***	.77***	.81***	.86***	.89***	1.00	
Procedural Justice	1.10	.23***	.62***	.40***	.44***	.75***	.78***	.87***	.83***	.96***	1.00

*Note: all Means = 0 for factor scores SD = standard deviation, statistical significance *** = p<0.01; ** = p<0.05; * = p<0.1.*

Table 7: Pearson correlation coefficients for latent variables using factor scores

Pearson correlation coefficients presented in Table 7 suggests Tech experience has a weak positive correlation to other variables, implying a weak linear relationship. All remaining latent variables have a moderate to strong positive correlation, implying a moderate to strong linear relationship. The three justice variables are very highly correlated.

4.3 Research question 1

RQ1: Does chatbot failure without adequate recovery strategies lead to a decrease in customer outcomes?

> **H_1:** Chatbot failure with no recovery leads to a decrease in customer outcomes

For this hypothesis testing, all failure scenarios were grouped together, and all recovery scenarios also grouped together. A t-test was performed to compare satisfaction, loyalty, and intent to re-use, before and after experiencing the service encounter for service failure and service recovery groups.

Table 8 depicts the t-test for service failure and recovery.

Variables	N	Mean	SD	DF	t-value	p value
Pre and Post Satisfaction for Failure group	22	0.1305	1.2713	21	0.48	0.6351
Pre and Post Satisfaction for Recovery group	80	-0.0362	1.0736	79	-0.3	0.7637
Pre and Post Loyalty for Failure group	46	-0.0479	0.6167	45	-0.53	0.6012
Pre and Post Loyalty for Recovery group	218	0.0101	0.7669	217	0.19	0.846
Pre and Post Re-use intent for Failure group	22	-0.0543	1.0252	21	-0.25	0.8062
Pre and Post Re-use intent for Recovery group	80	-0.1203	0.9435	79	-1.14	0.2574

Table 8: t-test for Service failure and Recovery

For the Service failure group, t-test results in Table 8 suggest there is no significant difference in loyalty, satisfaction, and intent to re-use when comparing means before and after the service failure encounter. **This contradicts the proposed hypothesis 1, confirming that chatbot failure with no recovery *does not* lead to a decrease in customer outcomes.**

4.4 Research question 2

RQ2: Do positive perceptions of recovery strategies improve customer outcomes over pre-failure levels?

4.4.1 *Research question 2A*

RQ2A: Do positive perceptions of recovery strategies improve *brand loyalty*?

H_1: Positive perceptions of recovery strategies *improve* brand loyalty.

For this hypothesis testing, a t-test was performed by grouping service failure scenarios and different recovery type scenarios to compare loyalty before and after the service encounter for each recovery type. Due to the limited sample across the 6 scenario types, similar recovery type groups were rationalised for this analysis. The ‘Digital guided- with procedures’ and ‘Digital guided – with recovery options’ groups were combined after confirming the same behaviour.

Table 9 includes the t-test results for pre and post loyalty by recovery type.

Recovery Strategy	*N*	*Mean*	*Std Dev*	*Minimum*	*Maximum*	*DF*	*t Value*	*Pr > |t|*
Failure - No Recovery	46	-0.05	0.62	-2.76	1.80	45	-0.53	0.601
Defer to Call Centre	40	0.03	0.63	-2.28	1.87	39	0.26	0.797
Digital Guided	94	0.15	0.79	-1.86	4.57	93	1.89	0.062*
Digital Navigation	37	-0.15	0.98	-5.27	0.71	36	-0.93	0.36
Defer to Online Agent	47	-0.17	0.58	-2.09	0.85	46	-1.98	0.054*

Table 9: T-test for Loyalty1 – Loyalty 2 by recovery strategy type

As shown in Table 9, recovery types ‘Defer to Online Agent’ and ‘Digital Guided’ have significantly different loyalty pre and post the service encounter scenario. All other recovery types do not show a significant difference in loyalty pre- and post the service encounter scenario. One-way ANOVA was performed for loyalty by recovery strategy type presented in Table 10 below.

Source	*DF*	*Type III SS*	*Mean Square*	*F Value*	*Pr > F*
Loyalty	1	0.16	0.16	0.58	0.446
*loyalty*Recovery Strategy*	4	2.24	0.56	2.06	0.086
Error(loyalty)	259	70.20	0.27		

Table 10: Analysis of Variance of loyalty by recovery strategy type

Table 10 suggests that loyalty is significantly different pre and post experiencing the service encounter scenarios (P = 0.086).

LS means were compared for loyalty pre and post service encounter to explore the variance by recovery type in Figure 4 below.

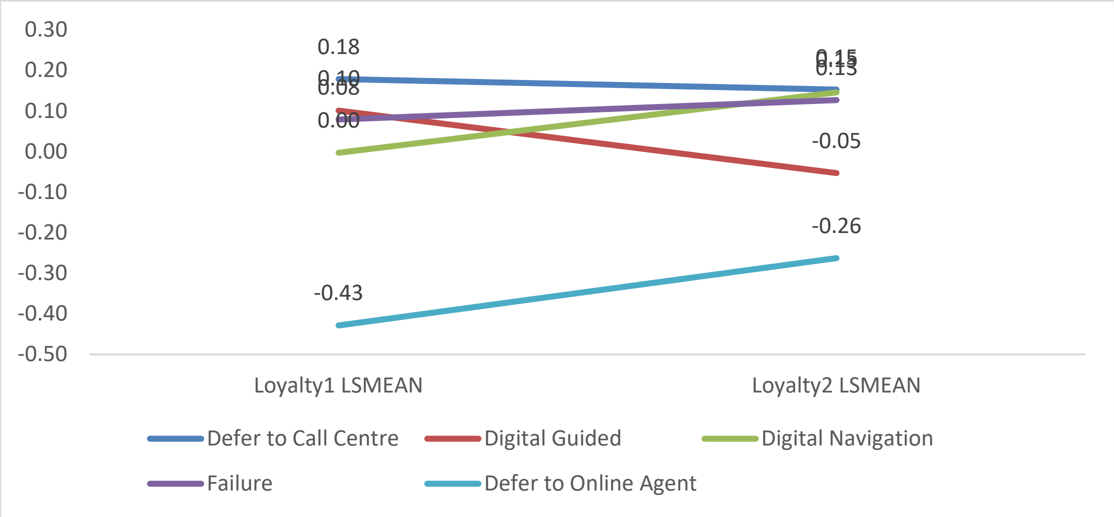


Figure 4: Pre and Post LS Means for Loyalty by Recovery Strategy type

As shown in figure 4, ‘Defer to Online Agent’ and ‘Digital Navigation’ recovery types show an upward trajectory of loyalty, however only ‘Defer to Online Agent’ is significantly different (p= 0.054) with 90% confidence level. **This finding supports hypothesis 2A, confirming that positive perceptions of recovery strategies improve brand loyalty.**

4.4.2 *Research question 2B*

RQ2B: Do positive perceptions of recovery strategies improve *intent to re-use* the chatbot service for future banking services?

> **H_1:** Positive perceptions of recovery strategies *increase* intent to re-use the chatbot service.

One way ANOVA and a t-test was performed to analyse ‘intent to re-use’ the chatbot, with the same groupings as previously done for loyalty testing. Table 11 present the t-test results for pre and post re-use by recovery type.

Recovery Strategy	***N***	***Mean***	***Std Dev***	***Minimum***	***Maximum***	***DF***	***t Value***	***Pr > |t|***
Failure - No Recovery	22	-0.0543	1.0252	-1.6583	2.1948	21	-0.25	0.806
Defer to Call Centre	15	-0.2615	1.0026	-2.9657	1.335	14	-1.01	0.330
Digital Guided	34	0.1674	1.0974	-2.7682	2.8683	33	0.89	0.380
Digital Navigation	11	-0.2865	0.5282	-1.4262	0.223	10	-1.8	0.102

Defer to Online Agent	20	-0.4121	0.6724	-1.6935	0.7298	19	-2.74	0.013**

Table 11: t-test for Re-use1 – Re-use 2 by recovery strategy type

Table 11 depicts ‘Defer to Online Agent’ has significantly different intent to re-use pre and post the service encounter scenario. All other recovery types do not show a significant difference in intent to re-use pre- and post the service encounter scenario.

Table 12 presents the ANOVA of intent to Re-use.

Source	*DF*	*Type III SS*	*Mean Square*	*F Value*	*Pr > F*
reuse	1	1.27	1.27	2.82	0.096
*reuse*Recovery Strategy*	4	2.60	0.65	1.44	0.225
Error(reuse)	97	43.64	0.45		

Table 12: Analysis of Variance of Intent to Re-use

Table 12 depicts that intent to re-use the chatbot is not significantly different pre and post experiencing the service encounter scenarios. Based on the t-test results ‘Defer to Online Agent’ is the only recovery type having a significant difference, therefore LS means were analysed further to confirm the hypothesis in Figure 5 below.

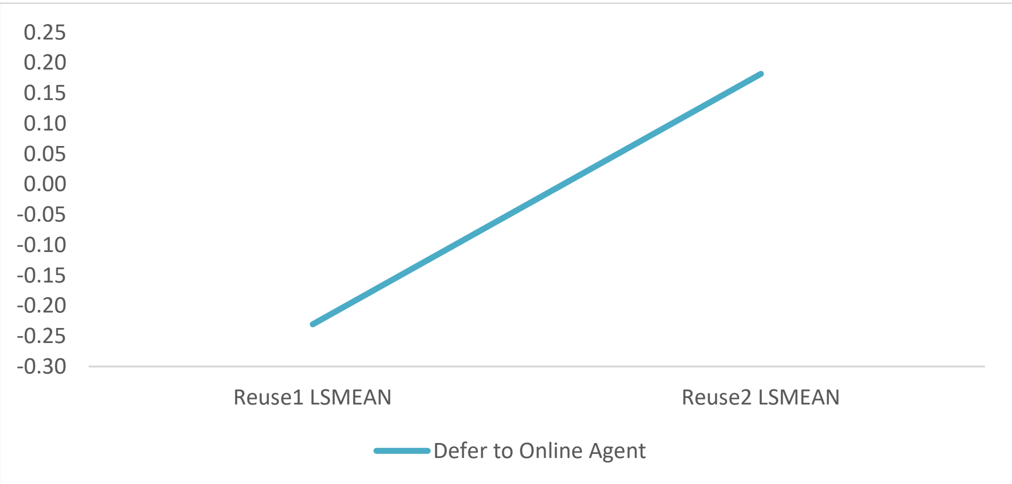


Figure 5: Pre and Post LS Means for Intent to Re-use by Recovery Strategy type

Figure 5 shows an upward trajectory for ‘Defer to Online Agent’ and ‘Digital Navigation’ and ‘Defer to Call Centre’ recovery types. Based on the t-test ‘Defer to Online Agent’ is significantly different (p= 0.013) with 90% confidence level. Digital Navigation’ and ‘Defer to Call Centre’ recovery types were not considered as the t-test suggests they are not significantly different.

Based on the significant upward trajectory for ‘Defer to Online Agent’ finding supports hypothesis 2b, this supports the hypothesis that positive perceptions of recovery strategies improve brand loyalty.

4.4.3 *Research question 2C*

RQ2C: Do positive perceptions of recovery strategies improve *satisfaction*?

H_1: Positive perceptions of recovery strategies *improve* satisfaction.

Table 13 below presents the t-test results for satisfaction by recovery strategy type.

Recovery Strategy	*N*	*Mean*	*Std Dev*	*Minimum*	*Maximum*	*DF*	*t Value*	*Pr > |t|*
Failure - No Recovery	22	0.1305	1.2713	-2.0111	3.4693	21	0.48	0.635
Defer to Call Centre	15	-0.3512	1.0747	-2.4056	1.0759	14	-1.27	0.226
Digital Guided	34	0.2517	1.301	-2.2148	4.2882	33	1.13	0.267
Digital Navigation	11	-0.3884	0.7781	-1.7587	0.7713	10	-1.66	0.129
Defer to Online Agent	20	-0.0957	0.6172	-1.1464	1.2168	19	-0.69	0.496

Table 13: t-test for Satisfaction1 – Satisfaction2 by recovery strategy type

Table 14 presents the AVOVA for satisfaction.

Source	*DF*	*Type III SS*	*Mean Square*	*F Value*	*Pr > F*
Satisfaction	1	0.36	0.36	0.59	0.444
*Satisfaction*Recovery Strategy*	4	3.11	0.78	1.27	0.289
Error (Satisfaction)	97	59.63	0.61		

Table 14: Analysis of Variance of Satisfaction

The t-test results depicted in table 13 shows that satisfaction is not significantly different pre and post service encounter for all recovery strategy types. This is confirmed by the ANOVA, in table 14 showing the relationship between recovery strategy type and satisfaction is not significant. Based on this, no further analysis was performed for satisfaction. **This confirms the null hypothesis, that positive perceptions of recovery strategies do not improve satisfaction.**

4.5 Research question 3

RQ3: Which recovery strategy is most likely to meet customer expectations best leading to most improved customer outcomes over pre-failure levels?

4.5.1 *Research question 3A*

RQ3A: Does handover to a human agent lead to a higher increase in customer outcomes over pre-failure levels than other recovery strategies?

This hypothesis was only tested for loyalty since it is the only customer outcome that was found to have more than one type of recovery strategy that is statistically significant for the comparison required for this hypothesis testing. The hypothesis was adjusted to specify outcomes in terms of loyalty as follows:

> **H_1:** Handover to a human agent *leads to* a higher increase in ***loyalty*** over pre-failure levels than other recovery strategies.

Referring to Figure 4, recovery types ‘Defer to Online Agent’ and ‘Digital Guided’ have significantly different loyalty pre and post the service encounter scenario. All other recovery types do not show a significant difference in loyalty pre- and post the service encounter scenario. Based on significance, ‘Defer to Online Agent’ and ‘Digital Guided’ were the only scenarios used to test this hypothesis, ‘Digital Guided’ being a non-human assisted recovery strategy type, and ‘Defer to Online Agent’ being a human-assisted recovery strategy type.

Figure 4 has been amended to reflect only recovery types of statistical significance for this hypothesis testing, depicted below in Figure 6.

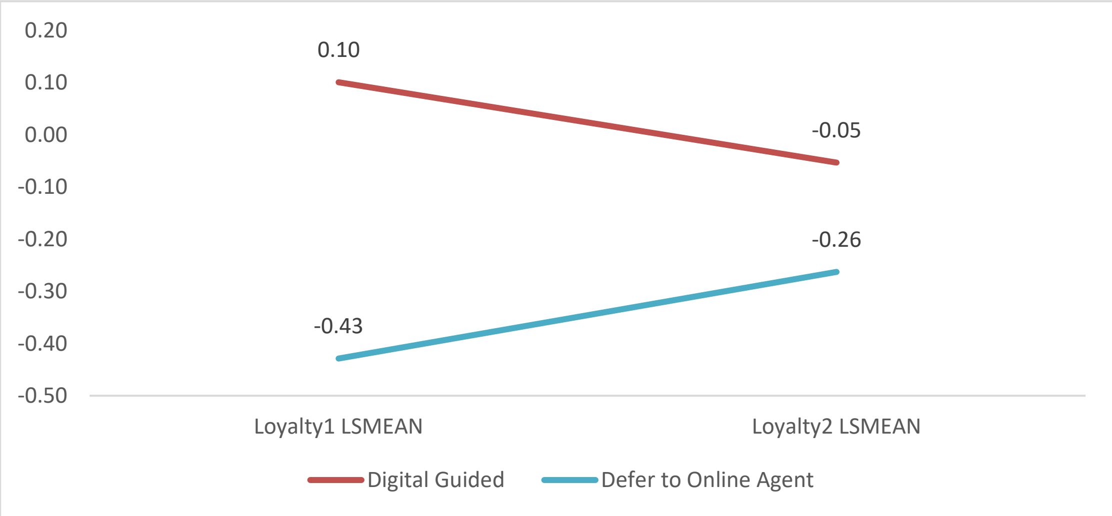


Figure 6: Pre and Post LS Means for Loyalty by Recovery Strategy type showing only statistically significant recovery types

As shown in figure 6, 'Defer to Online Agent' has an upward trajectory for loyalty, while 'Digital Guided' has a downward trajectory for loyalty. **This supports hypothesis 3A, confirming that handover to a online human- assisted agent leads to a higher increase in loyalty over pre-failure levels than other recovery strategies.**

This hypothesis does not hold true for handover to a human-assisted agent in the call centre scenario, as no significant difference was found. **This confirms that handover to a call centre agent does not lead to a higher increase in loyalty over pre-failure levels than other recovery strategies.**

4.5.1 *Research question 3B*

H3B: Does a sense of distributive justice restoration or equity rebalancing partly explain the change in customer outcomes following successful recovery strategies?

This hypothesis was only tested for loyalty since it is significantly different. The hypothesis was adjusted to specify outcomes in terms of loyalty as follows:

> **H_1:** A sense of distributive justice or equity rebalancing partly explains the change in **loyalty** following successful recovery strategies.

Analysis of Variance was performed using distributive justice as a mediating variable. Table 15 presents the ANOVA for loyalty with distributive justice as an interaction variable.

Source	***DF***	***Type III SS***	***Mean Square***	***F Value***	***Pr > F***
Loyalty	1	10.44	10.44	46.41	<.0001
*loyalty*Distributive Justice*	1	12.16	12.16	54.06	<.0001
*loyalty*Recovery Strategy*	4	1.95	0.49	2.17	0.0733
Error(loyalty)	258	58.04	0.22		

Table 15: Analysis of Variance of Loyalty with Distributive justice as an interaction variable

The ANOVA analysis depicted in Table 15 shows a significant relationship between loyalty and distributive justice ($p <.0001$), and a significant difference in loyalty by recovery type ($p=0.0733$) using 90% confidence level.

This finding supports the hypothesis, **confirming that a sense of distributive justice or equity rebalancing partly explains the increase in loyalty following successful recovery strategies.**

4.5.2 *Research question 3C*

RQ3C: Does a perception of interactional justice interact with the relationship between the availability of a human handover and improved customer outcomes?

This hypothesis was only tested for loyalty since it is the only customer outcome that is statistically significant. Handover to a call centre agent was also excluded based on previous findings. The hypothesis was adjusted to specify outcomes in terms of loyalty for an online human handover as follows:

> **H₁:** A perception of interactional justice interacts with the relationship between the availability of a human online agent handover and improved ***loyalty***.

Analysis of Variance was performed for loyalty using distributive justice as a mediating variable. Table 16 presents the ANONA of loyalty with interactional justice as an interaction variable.

Source	***DF***	***Type III SS***	***Mean Square***	***F Value***	***Pr > F***
loyalty	1	9.87	9.87	43.27	<.0001
*loyalty*Interactional Justice*	1	11.34	11.34	49.72	<.0001
*loyalty*Recovery Strategy*	4	2.60	0.65	2.85	0.0246
Error(loyalty)	258	58.86	0.23		

Table 16: Analysis of Variance of Loyalty with Interactional Justice as an interaction variable

The ANOVA analysis depicted in Table 16 shows a significant relationship between loyalty and interactional justice (p <.0001). and a significant difference in loyalty by recovery type (p=0.0246) using 90% confidence level.

This finding supports the hypothesis, **confirming that a perception of interactional justice interacts the relationship between the availability of a human online agent handover and improved *loyalty*.**

4.5.1 *Research question 3D*

RQ3D: Does a perception of procedural justice interact with the relationship between the service recovery type and improved customer outcomes?

This hypothesis was only tested for loyalty since it is the only customer outcome that is statistically significant. The hypothesis was adjusted to specify outcomes in terms of loyalty as follows:

> **H_1:** A perception of procedural justice *interacts with* the relationship between the service e recovery type and improved **loyalty**.

Analysis of Variance was performed for loyalty using procedural justice as a mediating variable in Table 17 below.

Source	***DF***	***Type III SS***	***Mean Square***	***F Value***	***Pr > F***
loyalty	1	9.29	9.29	40.28	<.0001
*loyalty*Procedural Justice*	1	10.68	10.68	46.31	<.0001
*loyalty*Recovery Strategy*	4	2.65	0.66	2.87	0.0235
Error(loyalty)	258	59.52	0.23		

Table 17: Analysis of Variance of Loyalty with Procedural Justice as interaction variable

The ANOVA analysis depicted in Table 17 shows a significant relationship between loyalty and interactional justice ($p <.0001$) and a significant difference in loyalty by recovery type ($p=0.0235$) using 90% confidence level.

This finding supports the hypothesis, **confirming that a perception of procedural justice interacts with the relationship between the service e recovery type and improved loyalty.**

4.5.1 *Research question 3E*

Are more digitally native customers more tolerant of alternative non-human digital interfaces as an initial recovery strategy rather than going straight to human agent, than would non digital natives?

> **H_1:** Digitally native customers are more tolerant of alternative non-human digital interfaces as an initial recovery strategy rather than going straight to human agent, than would non digital natives.

This hypothesis was tested only using loyalty as customer outcome based on previous findings. Analysis of Variance was performed for loyalty using Tech experience as an interaction variable.

Source	***DF***	***Type III SS***	***Mean Square***	***F Value***	***Pr > F***
loyalty	1	0.16	0.16	0.58	0.446
*loyalty*Tech_Experience*	1	0.45	0.45	1.65	0.200
*loyalty*Recovery Strategy*	4	2.10	0.53	1.94	0.104
Error(loyalty)	258	69.76	0.27		

Table 18: Analysis of Variance of Loyalty with Tech Experience as an interaction variable

The ANOVA analysis depicted in Table 18 shows a non- significant relationship between loyalty and tech experience (p=0.2). and there is no significant difference in loyalty by recovery type (p=0.104). This finding contradicts the hypothesis, **confirming that digitally native customers are not more tolerant of alternative non-human digital interfaces as an initial recovery strategy rather than going straight to human agent, than would non digital natives.**

No further analysis was performed based on tech experience variable due to no statistical significance.

4.5.2 *Research question 3F*

Are customers more tolerant of service failure when asking the chatbot complex questions, than they would when asking the chatbot fundamental financial services questions?

H_1: Customers *more tolerant* of service failure when asking the chatbot complex questions, than they would when asking the chatbot fundamental financial services questions.

This hypothesis was tested only using loyalty as customer outcome based on previous findings. ANOVA was performed for loyalty using service request **Salience** as an interaction variable in Table 19 below.

Source	***DF***	***Type III SS***	***Mean Square***	***F Value***	***Pr > F***
Loyalty	1	0.05	0.05	0.2	0.656
*loyalty*Salience*	1	0.01	0.01	0.03	0.855
*loyalty*Recovery Strategy*	4	2.18	0.54	2	0.095
Error(loyalty)	258	70.19	0.27		

Table 19: Analysis of Variance of Loyalty with Salience of Service Request as an interaction variable

The ANOVA analysis depicted in Table 19 shows a non- significant relationship between loyalty salience (p =0.855). There is a borderline significant difference in loyalty by recovery type (p=0.095) at 90% confidence level. This finding is not conclusive, hence the null hypothesis is assumed, **confirming that customers are not more tolerant of service failure when asking the chatbot complex questions, than they would when asking the chatbot fundamental financial services questions.**

No further analysis was performed based on salience due to inconclusive statistical significance.

4.6 Summary of results

The following findings have been confirmed based on statistical data analysis.

H1: Chatbot failure without adequate recovery strategies ***does not* lead to a decrease in customer outcomes**.

H2: Positive perceptions of recovery strategies ***improve brand loyalty and intent to re-use, however, does not improve satisfaction***.

H3a: Handover to a human online agent **leads** to a higher increase in loyalty over pre-failure levels than other recovery strategies?

H3b: A sense of distributive justice restoration or equity rebalancing ***partly explains the change in loyalty*** following successful recovery strategies.

H3c: A perception of interactional justice **interacts with** the relationship between the availability of a human online agent handover and improved customer outcomes.

H3d: A perception of procedural justice **interacts with** the relationship between the service recovery type and improved loyalty.

H3e: Digitally native customers **are not** more tolerant of alternative non-human digital interfaces as an initial recovery strategy rather than going straight to human handover, than would non digital natives?

H3f: Customers **are not** more tolerant of service failure when asking the chatbot salient questions, than they would when asking the chatbot fundamental financial services questions?

Based on these findings, **'Defer to Online Agent'** is the recovery type determined to **most likely to meet customer expectations** best leading to the most improved *loyalty* and intent to re-use over pre failure levels. **'**

CHAPTER 5. DISCUSSION OF RESULTS

5.1 Introduction

The purpose of this research was to understand the dynamics of chatbot service failure in financial services to implement effective recovery strategies that would lead to the most improved customer outcomes over pre-failure levels. This research focused on customer satisfaction, brand loyalty, and the intent to re-use the chatbot service post recovery, as these customer outcomes are pivotal to customer patronage.

This research is underpinned the expectation confirmation theory by Oliver (1980), Adams (1963) equity theory and justice theory, upon which extant literature on traditional service recovery with human agents is based. As technology has advanced, chatbots open the door to the use of artificial intelligence to simulate human conversation, shifting to technology-based servicing or hybrid thereof, as opposed to the traditional human-based servicing that is well researched (Ashktorab et al., 2019; Magnusson & Rånnerud, 2019).

5.2 Research Question 1 Discussion

The objective of the first research question was to determine whether chatbot failure without adequate recovery strategies leads to a decrease in customer outcomes?

The results suggest that chatbot failure without adequate recovery strategies ***does not*** lead to a decrease in customer outcomes.

In this study, 46 respondents experienced service failure scenarios with no recovery. No significant difference was found between satisfaction, loyalty, and intent to re-use pre- and post-service failure. This result suggests that chatbot service failure may not be considered as important as compared to service failure when assisted by a human, as observed in literature (Magnini et al., 2007; Tax et al., 1998). It is important to note that a direct comparison of technology service failure and human service failure was not directly tested in this study, and further research is required to confirm this emerging hypothesis.

Interpreting this finding in terms of expectation confirmation theory, this would imply that survey participants prior expectations of the chatbot service were met despite the service failure. This may be due to participants being unfamiliar of the capability of chatbots in service fulfilment and may change as the chatbot technology matures in South African financial services. The alternative rationale for this result is that the sample size was insufficient to detect minor variations in customer outcomes. Pairwise testing methods may be a more appropriate method of research to compare service failure and recovery strategy scenarios in future research.

5.3 Research Question 2 Discussion

The objective of the second research question was to determine whether positive perceptions of recovery strategies improve customer outcomes over pre-failure levels, with a specific focus on brand loyalty, customer satisfaction and intent to re-use the chatbot.

A study by Andreassen (2000) based on equity and expectation confirmation theories suggested that equity is impacted by the perceived performance of the service recovery, and proved that the perceived quality of service recovery positively impacts disconfirmation and satisfaction. Tax et al. (1998) refers to the double-deviation effect in service recovery for complaint handling, in which both initial service failure and recovery attempts fail.

In this study, the findings suggest that **positive perceptions of chatbot service recovery results in improved customer outcomes**, and conversely negative perceptions of service recovery results in reduced customer outcomes. This is observed by the 'Defer to Online Call Centre' recovery strategy significantly increasing loyalty, and conversely, the 'Digital Guide' recovery strategy significantly decreasing loyalty. Based on the findings of research question 1 that service failure does not significantly reduce customer outcomes below pre-failure levels, the negative impact observed for ‘Digital Guide’ recovery strategy may suggest a double deviation type effect, proving that finding recovery strategy suited to our South African market is pivotal to achieve the desired customer outcomes.

A similar observation was made for ‘Defer to Online Agent’ recovery strategy where an increase in intent to re-use the chatbot was observed. This single recovery type was an exception, as no significant difference in intent to re-use was observed for all other recovery type tested.

No significant difference was observed for satisfaction with recovery for all recovery strategies. This is most likely due to the limited responses for satisfaction and intent to re-use, as pre-measures for satisfaction and intent to re-use the chatbot were only measured for respondents who had previously used the existing Bank X chatbot (102 responses), as compared to a higher sample size for loyalty (264 responses).

5.4 Research Question 3 Discussion

5.4.1 *Most suitable recovery strategy type*

The third research question tests which recovery strategy is most likely to improve customer outcomes over pre failure levels. Overall, ‘**Defer to Online Agent’** was found to be the recovery strategy that is most likely to meet customer expectations and yielding to the most improved customer outcomes over pre-failure levels. This recovery type was found to be the only recovery strategy with a significant improvement in loyalty and intent to re-use the chatbot.

This observation suggests a hybrid approach to effective recovery strategies using human-assistance via a digital channel. This positive result was not observed for human assistance via telephonic channel, and therefore suggests a positive shift in digital transformation of the South African market.

‘Digital guided’ recovery type displayed a significant decrease in loyalty post recovery. This supports observations by Tax et al. (1998) that failed recovery attempt can be perceived as a double deviation, as both initial failure event and recovery attempt fail, resulting in lower customer outcomes than pre-failure levels.

This observation suggests that unassisted self-help solutions are not preferred by the South African market. Virtual human assistance is preferred over solutions offering support for self-help.

5.4.2 *Justice dimensions*

Hypothesis 3b, 3c and 3c tested the interaction effect of the justice dimensions, testing whether distributive, interactional and procedural fairness interact with the relationship between service recovery type and improved branch loyalty.

The results suggest that distributive, interactional and procedural justice play an interacting role on chatbot service recovery, as a significant relationship was observed for loyalty with each justice dimension. A significant difference in loyalty was observed with the justice dimensions as interacting variables. This result suggests that service recovery using chatbot technology is perceived in terms of justice theory, and is in line with the perception of justice and equity observed for human-assisted service recovery (Petzer et al., 2017; Yim et al., 2003). The increase in loyalty observed confirms a perceived restoration of equity and justice for the ‘Defer to online agent’ service recovery type.

Perceived interactional justice is evidently shown in the ‘Defer to Online agent’ recovery type in which there is a handover to a virtual human to assist with recovery. This stands out above the ‘Defer to Call Centre’ recovery type, despite it also being human assisted, as the same behaviour is not observed for this recovery type. This suggests that the perceived human intervention is observed as significant despite it being virtual assistance. This observation presents a positive opportunity for migration to digital channels from face to face or voice assisted channels.

The results also confirm that procedural justice **interacts with** the relationship between the service recovery type and improved customer outcomes. This implies that the service recovery process is perceived as fair.

A strong linear correlation was also observed among all justice dimension supporting the observation of all justice dimensions mediating the relationship between service recovery and loyalty. This concludes that equity and justice theory explain the increase in loyalty observed which chatbot service recovery.

5.4.1 *Prior Technology experience*

The South African digital divide is significantly larger than developed countries, supporting the need to determine whether digital natives are more tolerant of alternative non-digital interfaces an initial recovery strategy rather than going straight to human handover, than would non digital natives. The results suggest no difference in loyalty between digital natives non digital natives. This finding is important to understand how to cater for the digital divide in the South African market. The results suggests that customers with limited prior experience with technology do not require unique recovery strategies to compensate for their limited experience in technology. This finding simplifies the approach to service recovery across all customer types, and limits complexity.

5.4.1 *Salience of Service request type*

No significant difference was observed in customer outcomes for service requests of varied level of importance to customers. This test was limited in that self-reported salience was used to test this hypothesis. Further research is recommended to improve the validity of this finding.

CHAPTER 6. CONCLUSIONS & RECOMMENDATIONS

6.1 Introduction

Building on equity theory, expectation confirmation theory and justice theory, this research aimed to understand the impact of organisational recovery strategies on customer outcomes including satisfaction, customer loyalty and intent to re-use the chatbot service**.**

6.2 Conclusions research question 1

This study concludes that post experiencing a chatbot service failure, customer satisfaction, brand loyalty and intent to re-use, do not significantly differ from pre-failure levels. This observation suggests that participants prior expectations of the chatbot service were met despite the service failure. This may be due to participants being unfamiliar with the capability of chatbots in the nascent environment.

6.3 Conclusions regarding research question 2

Analysis by recovery type confirms that positive perceptions of recovery strategies improved customer loyalty and intent to re-use, however no difference was observed for satisfaction post recovery. This included a significant improvement in loyalty for “Defer to Online agent’ and a significant decline in loyalty for ‘Digital guided’ recovery type. The ‘Defer to Online Agent’ recovery strategy type showed a significant improvement in loyalty for intent to reuse variable, however this was not observed for all other recovery strategy types.

This result suggests that positive perceptions of recovery strategies can improve customer outcomes, and conversely negative perceptions of recovery strategies can have adverse effects on customer outcomes (Andreassen, 2000). This supports the view that identifying effective recovery strategies is pivotal to improving customer outcomes for customer retention and limit defective behaviour (Dwivedi et al., 2012).

6.4 Conclusions regarding research question 3

Defer to Online Agent’ was found to be the recovery strategy that is most likely to meet customer expectations, and best lead to the most improved customer outcomes over pre-failure levels.

Analysis of justice dimensions as mediating variables confirm a significant relationship between each justice dimension with loyalty, and a significant relationship between loyalty and recovery strategy type. This confirms that justice and equity theory apply to technology-based servicing via chatbot, much like for human-based servicing.

The significant improvement in loyalty observed for ‘Defer to Online agent’ suggests that interactional justice is perceived even for virtual human assistance. This is a positive sentiment supporting migration from face-to-face and voice-assisted channels to digital channels with virtual human support.

Analysis by tech experience also confirmed that digital natives are not more tolerant than non-digital natives. This finding supports the approach of having standardised effective recovery strategies agnostic of technology experience of the user. This limits the complexity in developing and maintaining different recovery strategies for varying levels of tech experience, as is indicative of our South African digital divide.

In terms of importance of service task, no difference is observed in customer outcomes with self-reported salience of service tasks. This finding is limited as the use of self-reported measures is not as effective as experiencing a simulation of varied task salience.

6.5 Recommendations

I recommend Bank X implement a recovery strategy for chatbot service failure that ‘Defers to a human online agent’.

The use of ‘Digital guided’ recovery strategy type was not preferred and should be avoided as a decrease in loyalty was observed, suggesting a double deviation (service failure and failed recovery) (Tax et al., 1998). This recovery strategy group included a self-service approach guiding the user on how to navigate to the appropriate mobile app screen to fulfil the service, as well as customers required to select their preferred recovery option. Both ‘Digital guided’ recovery types followed the same negative trajectory for loyalty.

‘Defer to call centre’ is also not recommended, as no significant difference was observed in customer outcomes, despite this being a human-assisted channel. The following plan is recommended for Bank X to implement the most effective recovery strategy identified:

Design chatbot for handover to virtual agent based on scenario prototype developed for this research	Service Design team - included User experience and user interface design
Develop chatbot handover and virtual agent communication interface via mobile app	App development team
Set-up virtual human agent resource team available for immediate chatbot handover	Operations Manager
Run pilot with internal staff or ringfenced user base	Program manager (coordination) Operations management
Retrospective assessment of pilot implementation	Program manager

Improvement of design and handover interface based on pilot retrospective analysis	Design team
Production in roll out to mobile app users in release	App development
Continuous analysis of chatbot handovers to virtual agents, and continuous improvement of app to grow knowledge base and logic to reduce service failures and handover volume	Operational management

6.6 Suggestions for further research

This research was performed using experimental scenarios combined with quantitative surveys using Likert scales. Each customer was only exposed to an individual sample of either one failure or one recovery strategy scenario. This research design approach limits the ability to make comparative conclusions. The quantitative research approach also limits verbatim feedback from participants in which additional learnings are achievable through qualitative methods.

I recommend further quantitative analysis using pairwise comparison experiments. This allows participants to make comparisons between at least two recovery strategies (Ashktorab et al., 2019).

Qualitative methods are also recommended to collect verbatim on recovery strategies, as well as to identify points of frustration experiences. This can provide valuable insights into the design of recovery strategies, including identification of tolerance points in conversation loops that are inherent to nascent chatbots.

REFERENCES

Abdulquadri, A., Mogaji, E., Kieu, T., & Nguyen, P. (2021). Digital transformation in financial services provision: A Nigerian perspective to the adoption of chatbot. *Journal of Enterprising Communities: People and Places in the Global Economy*.

Abe. (2016). *How Financial Chatbots Are Transforming Digital Banking*

Adams, J. S. (1963). Towards an understanding of inequity. *The Journal of Abnormal Social Psychology*, *67*(5), 422.

Adams, J. S. (1965). Inequity in social exchange. In *Advances in experimental social psychology* (Vol. 2, pp. 267-299). Elsevier.

Andreassen, T. (2000). Antecedents to satisfaction with service recovery. *J European Journal of Marketing*.

Ashktorab, Z., Jain, M., Liao, Q., & Weisz, J. (2019). Resilient chatbots: Repair strategy preferences for conversational breakdowns. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems,

Bobâlcă, C., Gătej, C., & Ciobanu, O. (2012). Developing a scale to measure customer loyalty. *3*, 623-628.

Boshoff, C. (1999). RECOVSAT: An instrument to measure satisfaction with transaction-specific service recovery. *Journal of service research*, *1*(3), 236-249.

Buse, Ç., & Hassan, A. (2019). Bibliometric Profile of Studies Related with Equity Theory. *Journal of Tourismology*, *5*(1), 17-29.

Cai, R., & Qu, H. (2018). Customers' perceived justice, emotions, direct and indirect reactions to service recovery: Moderating effects of recovery efforts. *Journal of Hospitality Marketing*, *27*(3), 323-345.

Camarate, J., & C, M. (2018). *A marketplace without Boundaries 2.0 & Digital disruption in the South African banking sector* [Publication].

Corea, C., Delfmann, P., & Nagel, S. (2020). Towards Intelligent Chatbots for Customer Care-Practice-Based Requirements for a Research Agenda. Proceedings of the 53rd Hawaii International Conference on System Sciences,

del Río-Lanza, A. B., Vázquez-Casielles, R., & Díaz-Martín, A. (2009). Satisfaction with service recovery: Perceived justice and emotional responses. *Journal of business research*, *62*(8), 775-781.

Diederich, S., Lembcke, T.-B., Brendel, A. B., & Kolbe, L. M. J. A. T. o. H.-C. I. (2021). Understanding the impact that response failure has on how users perceive anthropomorphic conversational service agents: Insights from an online experiment. *13*(1), 82-103.

Dwivedi, Y., Wade, M., & Schneberger, S. (2012). *Information Systems Theory, Explaining and Predicting Our Digital Society* (Vol. Vol.1). Springer.

Eren, B. A. (2021). Determinants of customer satisfaction in chatbot use: evidence from a banking application in Turkey. *International Journal of bank marketing*.

Ilias, A., Abd Razak, M., & Abd Razak, S. (2014). The intention to re-use the internet financial reporting (IFR) in Malaysia. *Review of Integrative Business Economic Research*, *3*(1), 337-370.

Jasso, G., Törnblom, K., & Sabbagh, C. (2016). Distributive justice. In *Handbook of social justice theory and research* (pp. 201-218). Springer.

Jung, N. Y., Seock, Y.-K., & Services, C. (2017). Effect of service recovery on customers' perceived justice, satisfaction, and word-of-mouth intentions on online shopping websites. *Journal of Retailing*, *37*, 23-30.

Kelley, S. W., & Davis, M. A. (1994). Antecedents to customer expectations for service recovery. *Journal of the academy of Marketing Science*, *22*(1), 52-61.

Lai, M.-C., & Chou, F. (2015). The relationships among involvement level, service failure, service recovery disconfirmation and customer lifetime value. *Journal of Economics, Business Management*

3(4), 452-457.

Lii, Y., Ding, M., & Lin, C. (2018). Fair or unfair: The moderating effect of sustainable CSR practices on anticipatory justice following service failure recovery. *Journal of Sustainability*, *10*(12), 4548.

Magnini, V. P., Ford, J. B., Markowski, E. P., & Honeycutt, E. D. (2007). The service recovery paradox: justifiable theory or smoldering myth? *Journal of Services Marketing*

Magnusson, P., & Rånnerud, A. (2019). *Chatbots in service recovery: Crackpot or Jackpot?* Handelshøyskolen BI].

Mattila, A. S. (1999). An examination of factors affecting service recovery in a restaurant setting. *Journal of Hospitality Tourism Research*, *23*(3), 284-298.

Maxham, J. I. (2001). Service recovery's influence on consumer satisfaction, positive word-of-mouth, and purchase intentions. *Journal of business research*, *54*(1), 11-24.

Maxham, J. I., & Netemeyer, R. (2003). Firms reap what they sow: the effects of shared values and perceived organizational justice on customers' evaluations of complaint handling. *Journal of Marketing*, *67*(1), 46-62.

Mogaji, E., Abdulquadri, A., Kieu, T., & Nguyen, N. P. (2020). Digital Transformation in Financial Services Provision: A Nigerian Perspective to the Adoption of Chatbot. *Journal of Enterprising Communities: People*.

Oliver, R. (1980). A cognitive model of the antecedents and consequences of satisfaction decisions. *17*(4), 460-469.

Oliver, R. (1999). Whence consumer loyalty? *Journal of Marketing*, *63*(4_suppl1), 33-44.

Oliver, R., & Swan, J. (1989). Consumer perceptions of interpersonal equity and satisfaction in transactions: a field survey approach. *53*(2), 21-35.

Parasuraman, A., Berry, L., & Zeithaml, V. (1991). Understanding customer expectations of service. *Sloan management review*, *32*(3), 39-48.

Petzer, D., De Meyer-Heydenrych, C., & Svensson, G. (2017). Perceived justice, service satisfaction and behavior intentions following service recovery efforts in a South African retail banking context. *International Journal of bank marketing*.

Poser, M., Singh, S., & Bittner, E. (2021). Hybrid Service Recovery: Design for Seamless Inquiry Handovers between Conversational Agents and Human Service Agents. Proceedings of the 54th Hawaii International Conference on System Sciences,

Rasooli, A., Zandi, H., & DeLuca, C. (2019). Conceptualising fairness in classroom assessment: exploring the value of organisational justice theory. *Journal of Assessment in Educatnio: Principle, Policy Practice*, *26*(5), 584-611.

Review, M. T. (2016). The Biggest Technology Failures of 2016. https://www.cnbc.com/2016/12/28/mit-tech-review-lists-2016s-biggest-tech-failures.html

Ryan, J. C. (2016). Old knowledge for new impacts: Equity theory and workforce nationalization. *Journal of business research*, *69*(5), 1587-1592.

Schneider, B., & Bowen, D. (1999). Understanding customer delight and outrage. *Sloan management review*, *41*(1), 35-45.

Schuetzler, R. M., Giboney, J. S., Grimes, G. M., & Nunamaker Jr, J. F. (2018). The influence of conversational agent embodiment and conversational relevance on socially desirable responding. *Decision Support Systems*, *114*, 94-102.

Severt, D. E. (2002). *The customer's path to loyalty: a partial test of the relationships of prior experience, justice, and customer satisfaction*. Virginia Polytechnic Institute and State University.

Singh, A., Ramasubramanian, K., & Shivam, S. (2019). Building an Enterprise Chatbot.

Smith, A. K., & Bolton, R. N. (1998). An experimental investigation of customer reactions to service failure and recovery encounters: paradox or peril? *Journal of service research*, *1*(1), 65-81.

Sudzina, F. (2015). Do gender and personality traits (BFI-10) influence self-perceived tech savviness. Proceedings of the 18th International Conference on Information Technology for Practice 2015,

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International journal of medical education*, *2*, 53.

Tax, S., & Brown, S. (1998). Recovering and Learning from Service Failure. *Sloan management review*.

Tax, S., Brown, S., & Chandrashekaran, M. (1998). Customer evaluations of service complaint experiences: implications for relationship marketing. *Journal of Marketing*, *62*(2), 60-76.

Tech2. (2022). Facebook researchers shut down AI Bots That Started Speaking in a Language Unintelligable to Humans. https://www.firstpost.com/tech/news-analysis/facebook-researchers-shut-down-ai-bots-that-started-speaking-in-a-language-unintelligible-to-humans-3876197.html

Wirtz, J., & Mattila, A. (2004). Consumer responses to compensation, speed of recovery and apology after a service failure. *International Journal of Service Industry Management*.

Yani-de-Soriano, M., Hanel, P. H., Vazquez-Carrasco, R., Cambra-Fierro, J., Wilson, A., & Centeno, E. (2019). Investigating the role of customers' perceptions of employee effort and justice in service recovery: a cross-cultural perspective. *European Journal of Marketing*.

Yim, C. K. B., Gu, F. F., Chan, K. W., & David, K. (2003). Justice-based service recovery expectations: measurement and antecedents. *Journal of Consumer Satisfaction, Dissatisfaction*

Complaining Behavior, *16*, 36.

Zumstein, D., & Hundertmark, S. (2017). Chatbots An interactive technology for Personalised communication, transactions ans services. *IADIS International Journal on WWW/Internet*, *15*(1).

APPENDIX A: Pre-survey constructs

Dimension	Scale item							
Technology experience	I consider myself to be tech savvy (skilled with technology.	SD	D	N	A	SA		
	People consider me to be tech savvy (skilled with technology).	SD	D	N	A	SA		
	I consider myself to be an advanced technology user	SD	D	N	A	SA		
Prior use of Bank X's chatbot	I have used the Bank X chatbot before and have attempted to ask the chatbot a question.	Yes	No	Not sure				
Intent to re-use	Based on my experience, I would use the Bank X chatbot again soon.	SD	D	N	A	SA		
	Based on my experience, I would use the Bank X chatbot frequently in future.	SD	D	N	A	SA		
Customer satisfaction	My service request was resolved to satisfaction.	TD	SD	D	N	A	SA	TA
	I am happy with the way my request was handled.	TD	SD	D	N	A	SA	TA
	I am happy with the service I experienced.	TD	SD	D	N	A	SA	TA
Affective Loyalty	I am pleased to use Bank X instead of other financial services brands.	TD	SD	D	N	A	SA	TA
	I feel more attached to Bank X than other financial services brands.	TD	SD	D	N	A	SA	TA
Conative Loyalty	I intend to use the Bank X brand in future.	TD	SD	D	N	A	SA	TA
	I intend to use other products from Bank X.	TD	SD	D	N	A	SA	TA
Action Loyalty	I recommend Bank X to those who ask my advice.	TD	SD	D	N	A	SA	TA
	I consider Bank X my first choice when buying financial products and services.	TD	SD	D	N	A	SA	TA
	I say positive things about Bank X to other persons.							

Table 20: Pre-Survey scale items

APPENDIX B: Experiment scenario designs

Below are illustrations of the eighteen experimental chatbot scenario designs used in data collection, labelled from scenario #0 to #17.

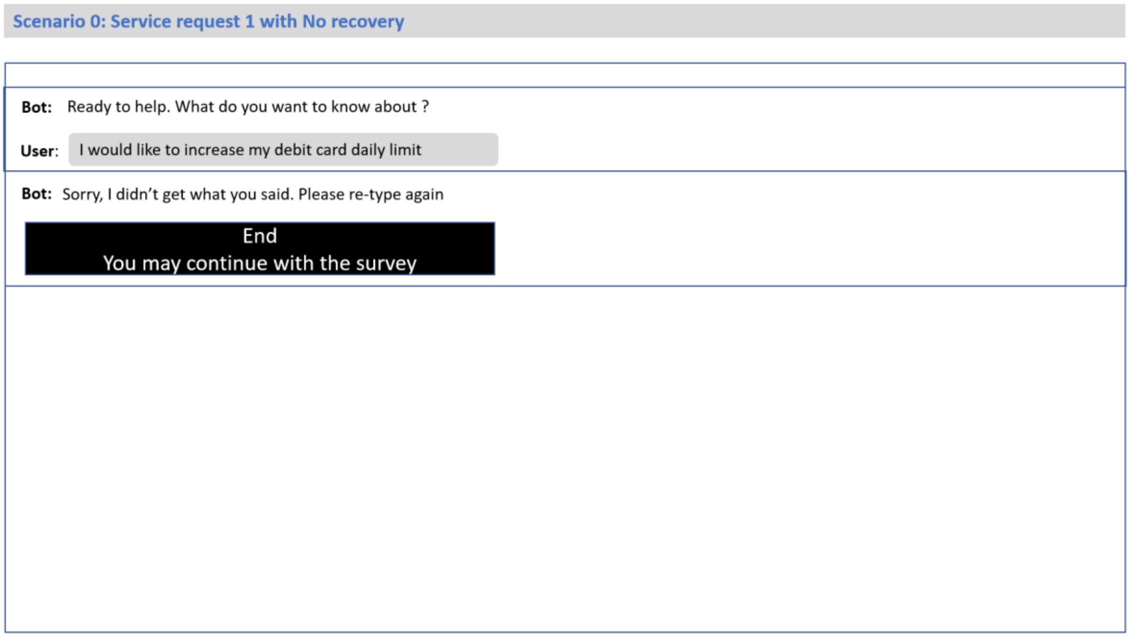


Figure 7: Chatbot scenario #0: Service request 1 with no recovery

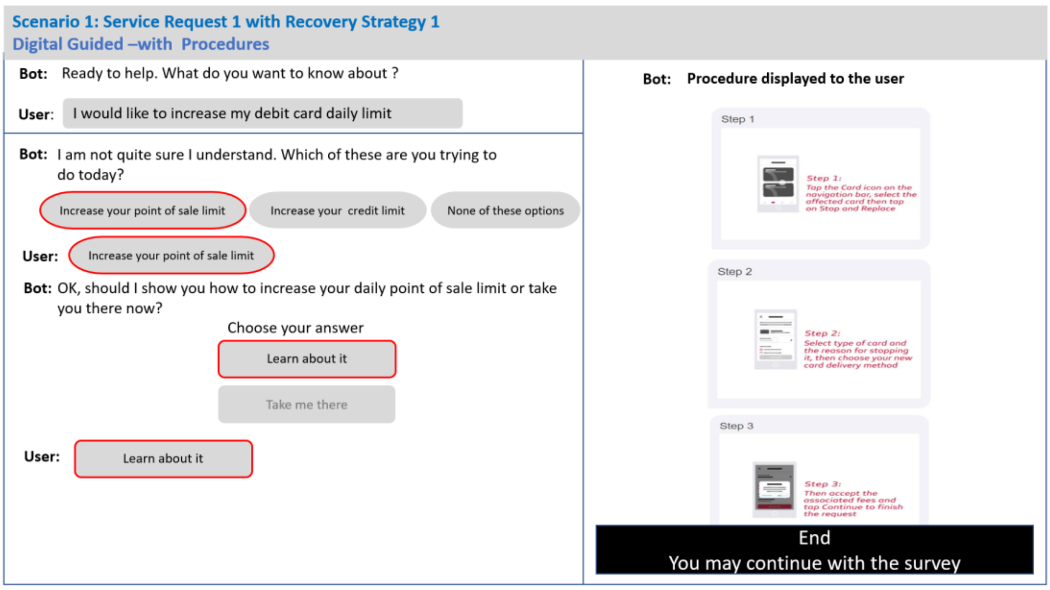


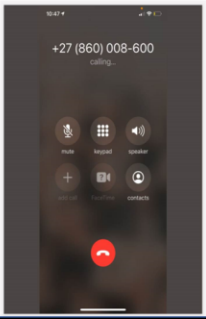
Figure 8: Chatbot scenario #1: Service request 1 recovery strategy 1

Scenario 2: Service Request 1 with Recovery Strategy 2
Defer to Human Call centre agent

Bot: Ready to help. What do you want to know about ?

User: I would like to increase my debit card daily limit

Bot: Sorry, I don't understand. For this request, I will need to transfer you to a human call centre agent who will address your request shortly.
Defer to Human Call Centre Agent - voice recorded call simulation played with call screen below displayed



End
You may continue with the survey

Figure 9: Chatbot scenario #2: Service request 1 recovery strategy 2

Scenario 3: Service Request 1 with Recovery Strategy 3
Defer to Human Online Agent

Bot: Ready to help. What do you want to know about ?

User: I would like to increase my debit card daily limit

Bot: Sorry, I don't understand. For this request, I will need to transfer you to a human call centre agent who will address your request shortly.
Defer to Human Call Centre Agent - voice recorded call simulation played with call screen below displayed

Online Agent: Good day, my name is Sipho. What can I help you with?

User: I would like to increase my debit card daily limit

Online Agent: I can direct you to change your card limit on your mobile App, select "Menu" in the bottom of your screen, and then select "Cards". Click on the card you want to change and select "Card transaction limits". Here you can change your Point of Sale transaction limit or ATM withdrawal limit, and then click save. Once this is done your new limits will take effect immediately and you will be able to purchase at a point of sale for the updated value. Is there anything else I can assist you with.

User: No further requests, thank you for your help

Online Agent: Thank you, Goodbye, and have a good day further.

End
You may continue with the survey

Figure 10: Chatbot scenario #3: Service request 1 recovery strategy 3

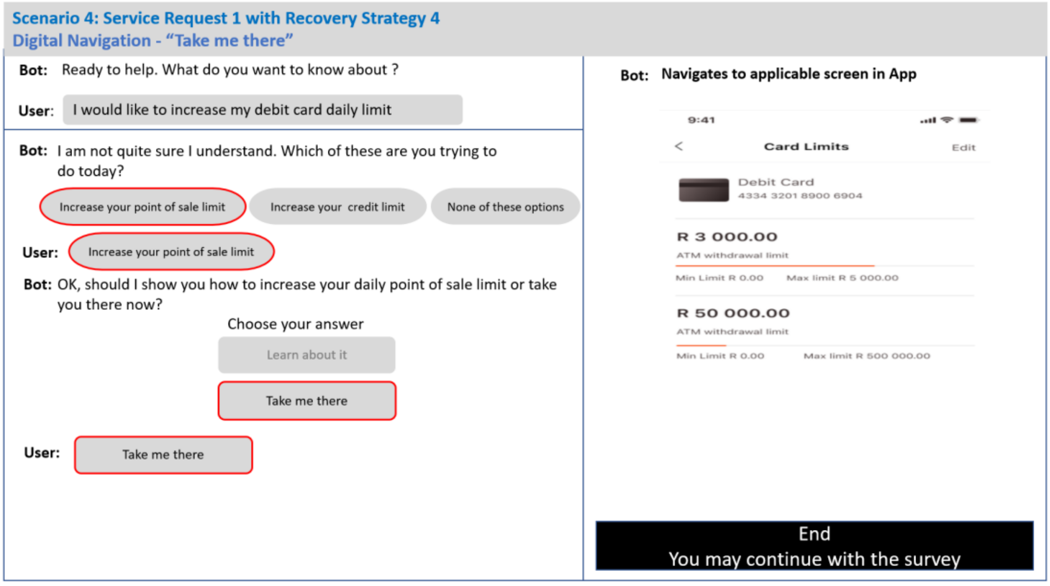


Figure 11: Chatbot scenario #4: Service request 1 recovery strategy 4

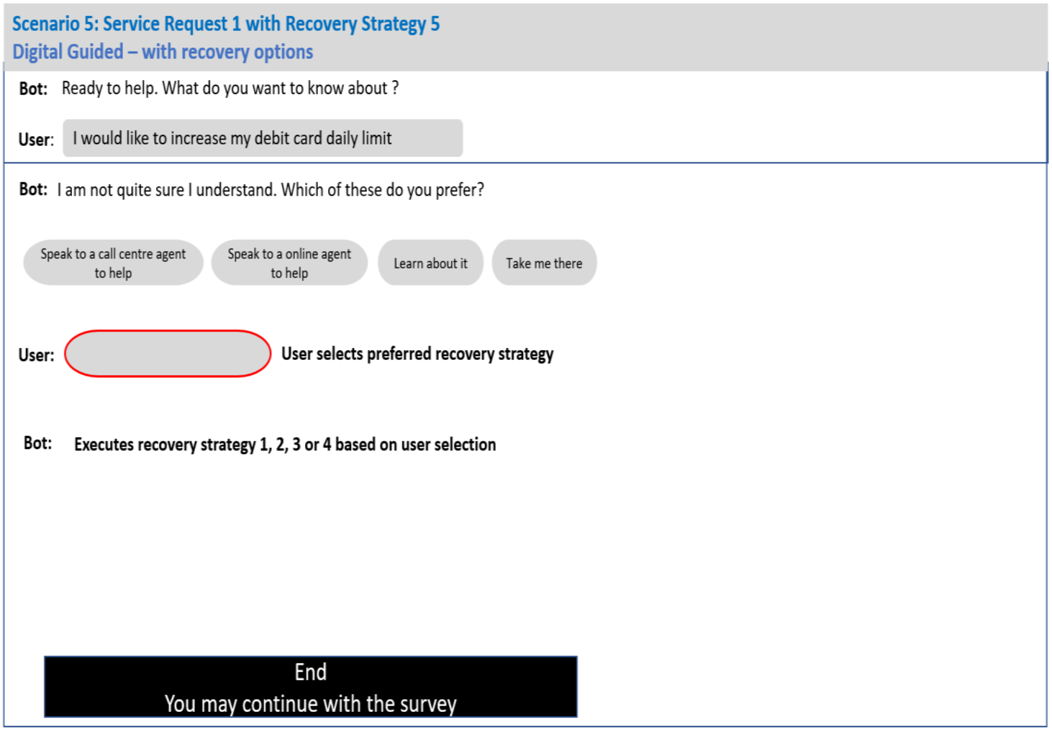


Figure 12: Chatbot scenario #5: Service request 1 recovery strategy 5

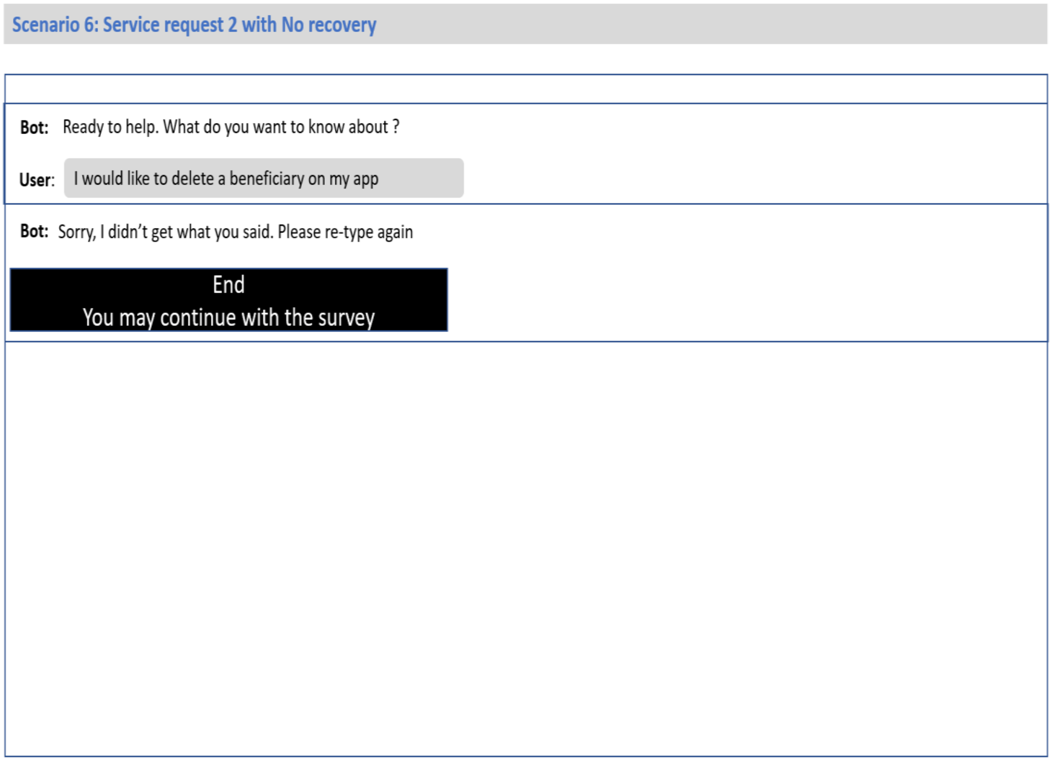


Figure 13: Chatbot scenario #6: Service request 2 with no recovery

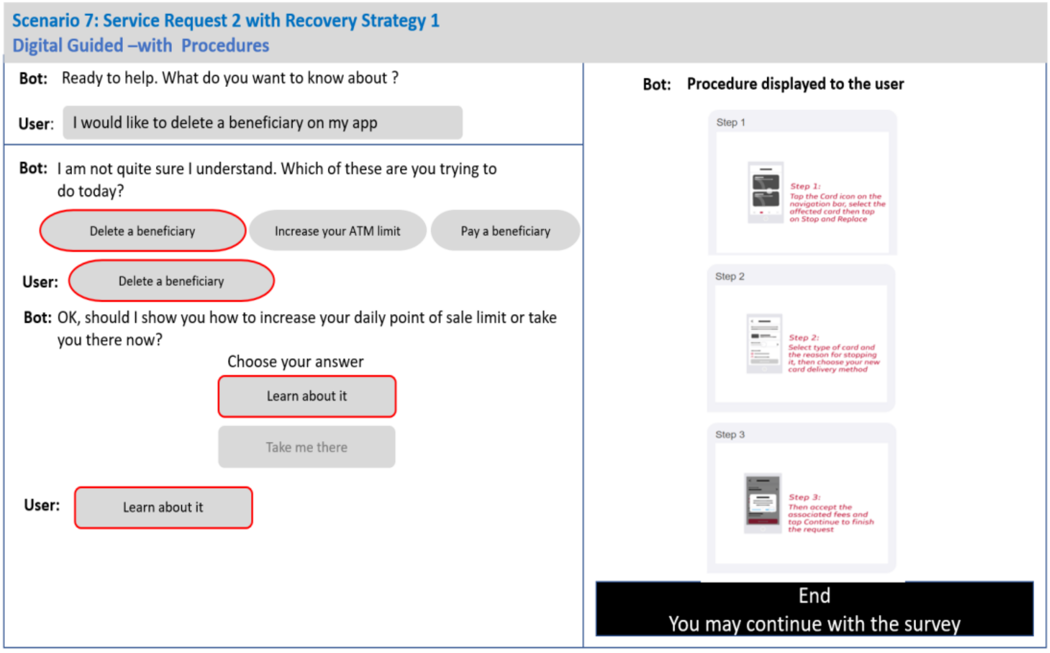


Figure 14: Chatbot scenario #7: Service request 2 recovery strategy 1

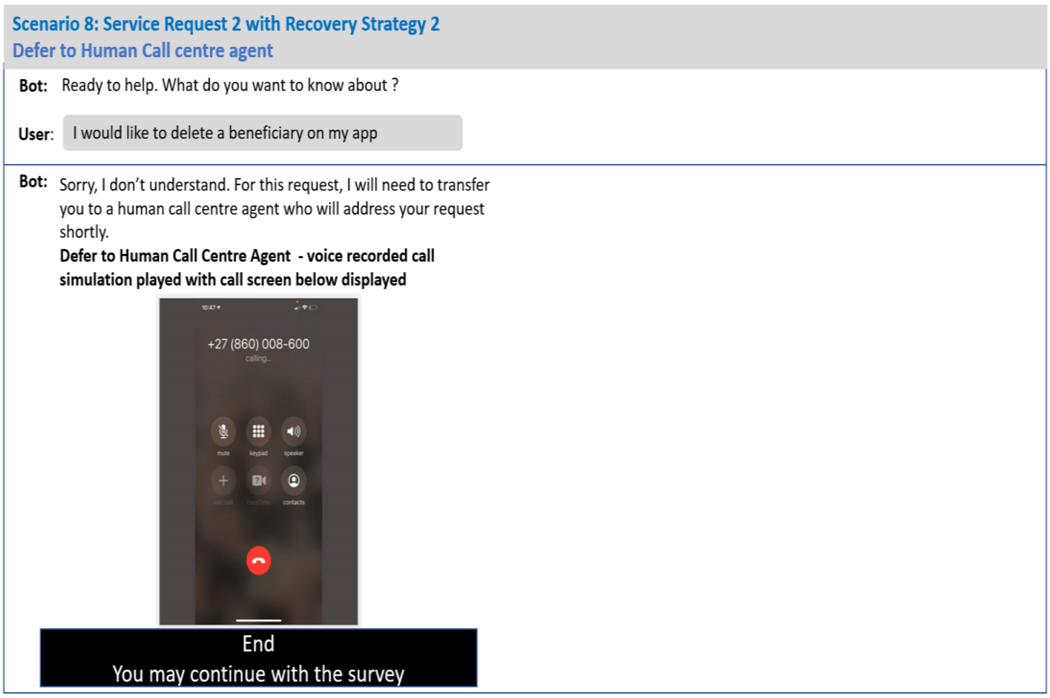


Figure 15: Chatbot scenario #8: Service request 2 recovery strategy 2

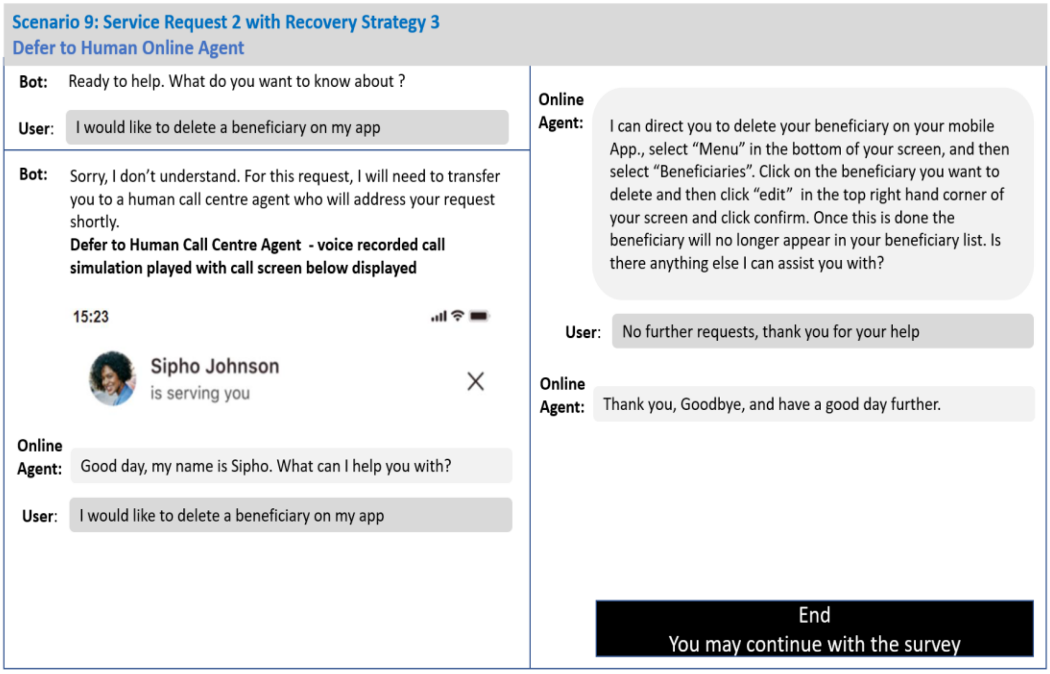


Figure 16: Chatbot scenario #9: Service request 2 recovery strategy 3

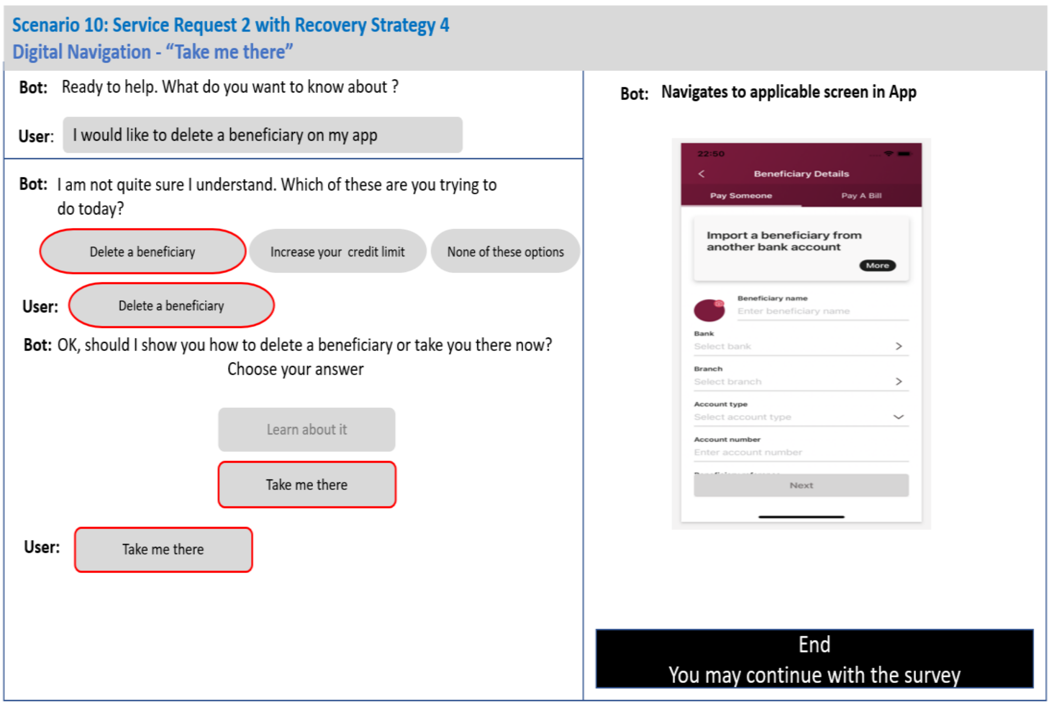


Figure 17: Chatbot scenario #10: Service request 2 recovery strategy 4

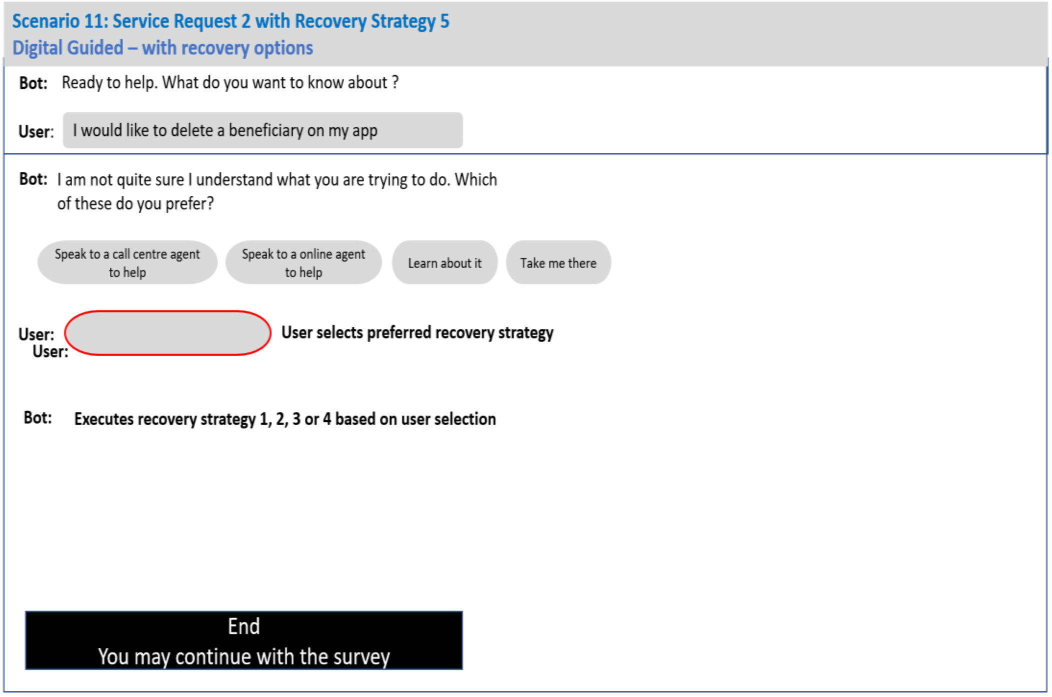


Figure 18: Chatbot scenario #11: Service request 2 recovery strategy 5

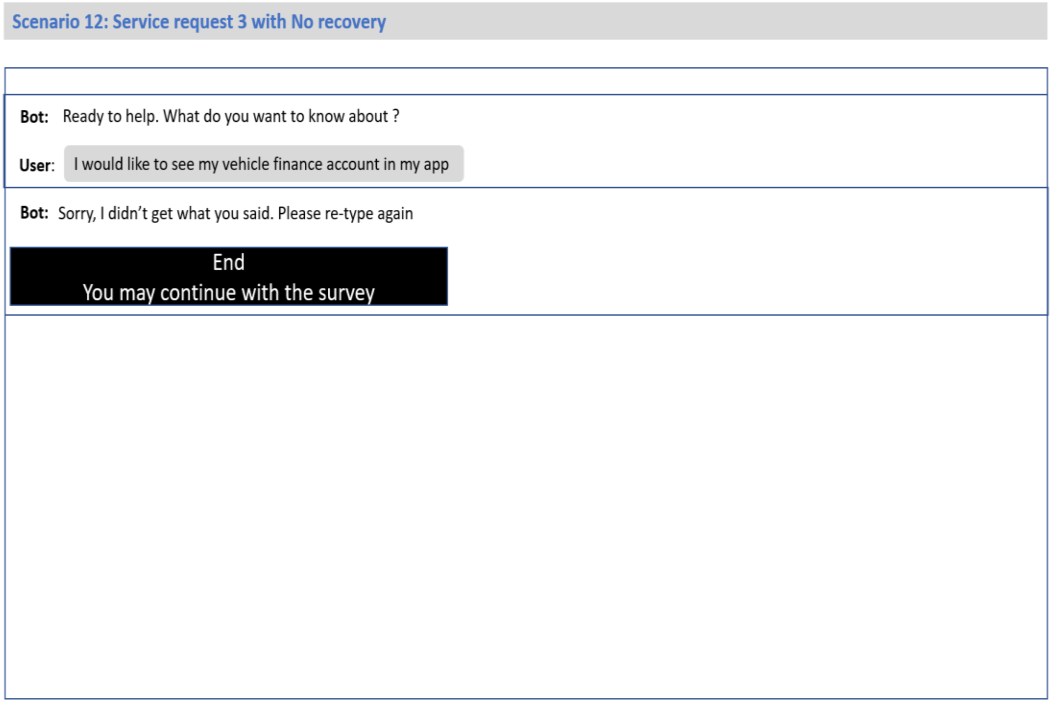


Figure 19: Chatbot scenario #12: Service request 3 with no recovery

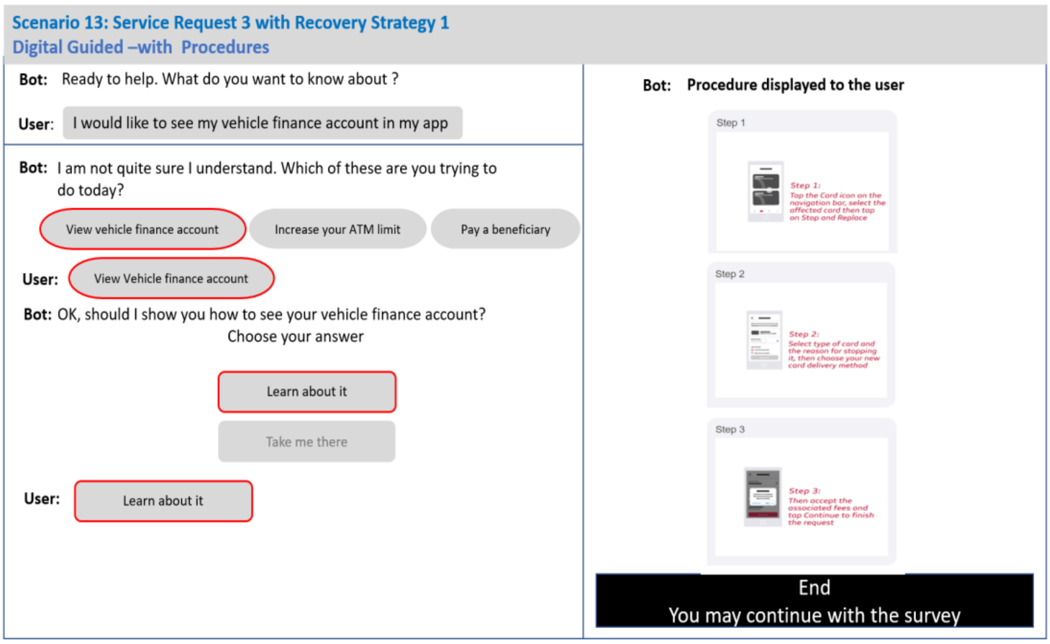


Figure 20: Chatbot scenario #13: Service request 3 recovery strategy 1

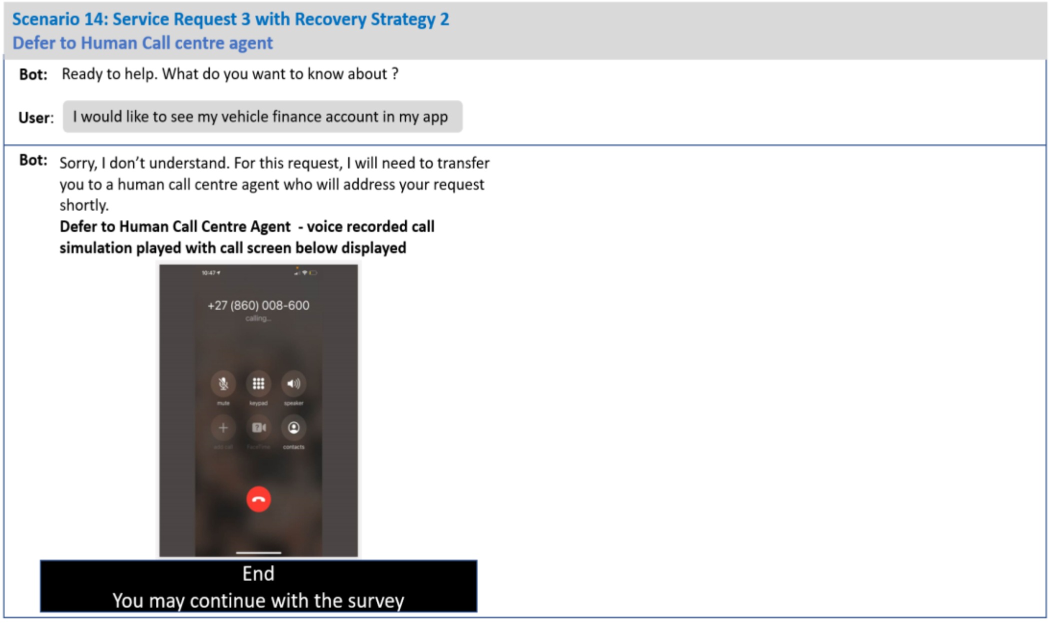


Figure 21: Chatbot scenario #14: Service request 3 recovery strategy

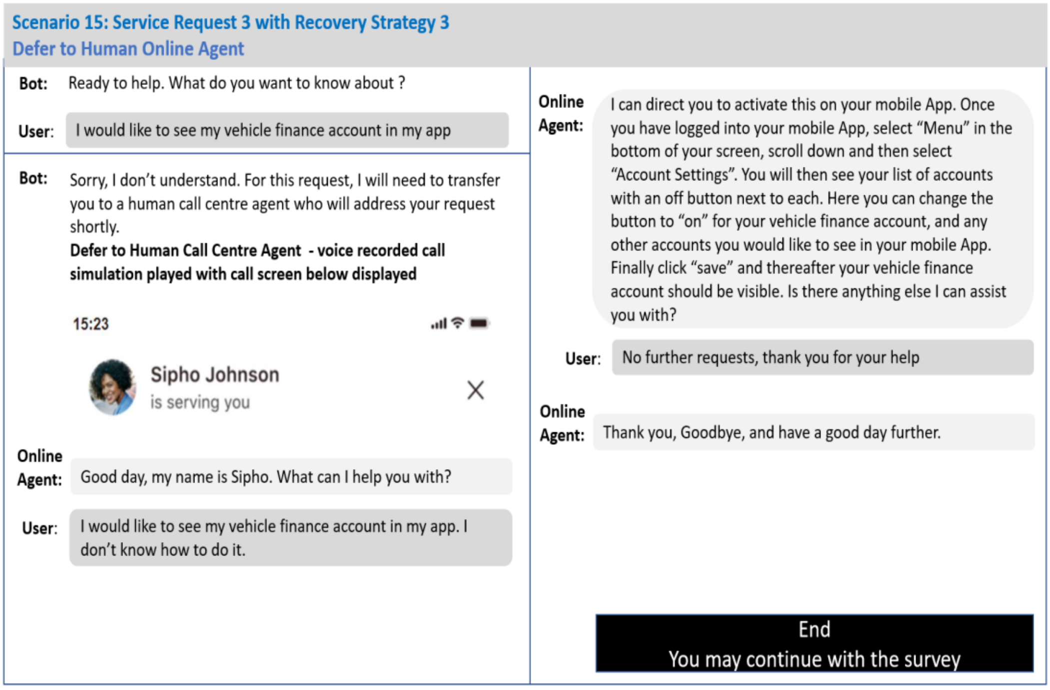


Figure 22: Chatbot scenario #15: Service request 3 recovery strategy 3

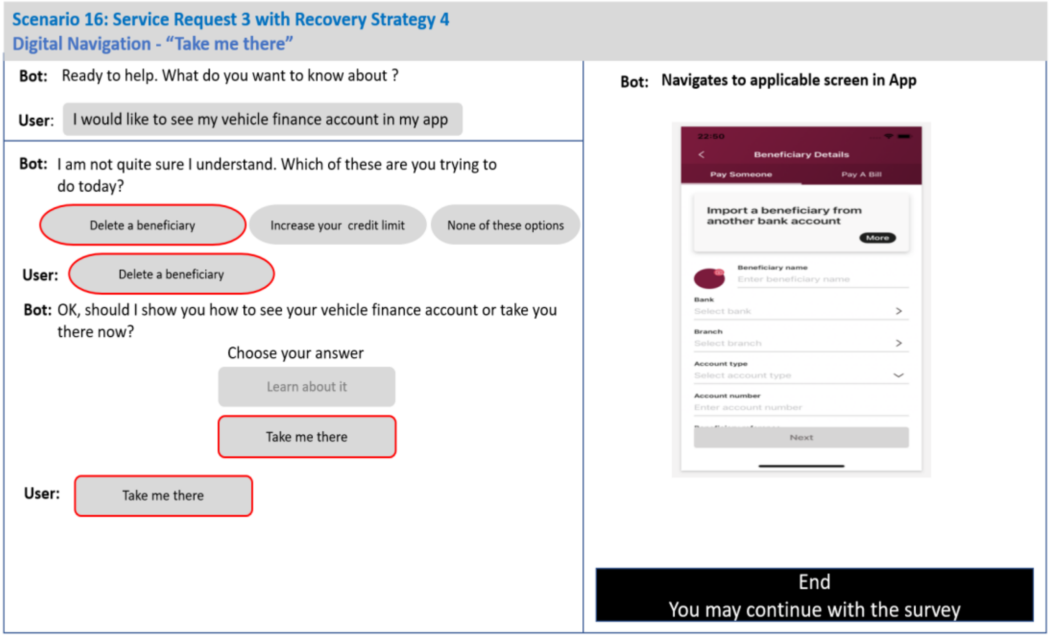


Figure 23: Chatbot scenario #16: Service request 3 recovery strategy 4

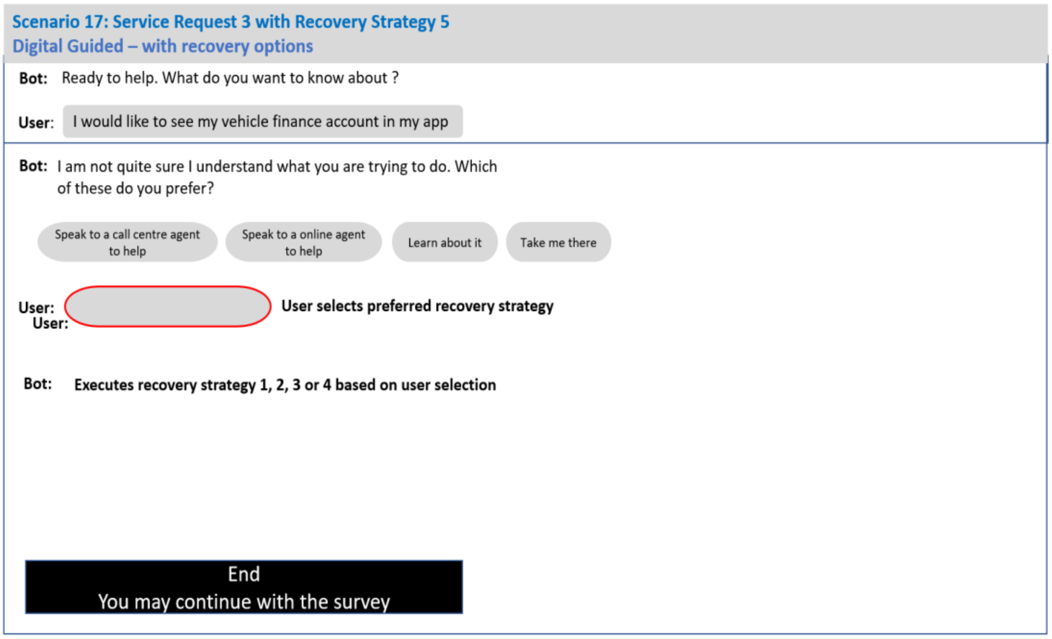


Figure 24: Chatbot scenario #17: Service request 3 recovery strategy 5

APPENDIX C: Post-survey constructs

Dimension	Scale item							
Based on this service encounter,								
Intent to re-use	I would use the Bank X chatbot again soon.	SD	D	N	A	SA		
	I would use the Bank X chatbot frequently in future.	SD	D	N	A	SA		
Customer Satisfaction	My service request was resolved to satisfaction.	TD	SD	D	N	A	SA	TA
	I am happy with the way my request was handled.	TD	SD	D	N	A	SA	TA
	I am happy with the service I experienced.	TD	SD	D	N	A	SA	TA
Affective Loyalty	I am pleased to use Bank X instead of other financial services brands.	TD	SD	D	N	A	SA	TA
	I feel more attached to Bank X than other financial services brands.	TD	SD	D	N	A	SA	TA
Conative Loyalty	I intend to use the Bank X brand in future.	TD	SD	D	N	A	SA	TA
	I intend to use other products from Bank X.	TD	SD	D	N	A	SA	TA
Action Loyalty	I recommend Bank X to those who ask my advice.	TD	SD	D	N	A	SA	TA
	I consider Bank X my first choice when buying financial products and services.	TD	SD	D	N	A	SA	TA
	I say positive things about Bank X to other persons.	TD	SD	D	N	A	SA	TA
Distributive Justice	I got the outcome I expected.	TD	SD	D	N	A	SA	TA
	I got the outcome I needed.	TD	SD	D	N	A	SA	TA
Interactional Justice	Bank X showed a real interest in my service request.	TD	SD	D	N	A	SA	TA

	Bank X treated me in a courteous manner when solving my service request.	TD	SD	D	N	A	SA	TA
	Bank X did everything possible to solve my service request.	TD	SD	D	N	A	SA	TA
Procedural Justice	My service request was resolved in the right way.	TD	SD	D	N	A	SA	TA
	Bank X has good policies and practices for dealing with my service requests.	TD	SD	D	N	A	SA	TA
	Despite the trouble experienced with the chatbot, Bank X was able to respond adequately.	TD	SD	D	N	A	SA	TA
	Bank X tried to solve my service request as quickly as possible.	TD	SD	D	N	A	SA	TA
	Bank X proved flexible in solving my service request.	TD	SD	D	N	A	SA	TA
Service request Salience	Rate the level of importance of the following service requests by importance to you							
	Increase my card point of sale limit in my mobile banking app	Not important	Less important	So so	Important	Very important		
	Assist me to delete a beneficiary in my mobile app	Not important	Less important	So so	Important	Very important		
	See my vehicle finance account in my mobile app							

Table 21: Post Survey scale items

APPENDIX D: Survey introductory letter

To whom it may concern

Bank X is continuously seeking ways to improve its customer service offering to you as a valued client. To this end, we have commissioned KLA, an independent external research provider, to conduct research participant recruitment services for the Chatbot initiative on our behalf in January 2022 and recruit people like yourself to complete the survey in question. KLA is a reputable, registered research company that subscribes to International Code on Market and Social Research as governed by ESOMAR.

A representative of KLA has contacted you to arrange for you to complete the online survey. Please be assured that KLA will treat all information as strictly confidential and it will not be used for any purpose other than research. The research results will only be used by Bank X at an aggregated level, and you will not be identified in a personal capacity in the final report. Please also note that you won’t be asked to divulge any personal, internet or bank account details at any point in the survey.

The competition of the survey is linked to a competition amounting to R10, 000 worth of vouchers. 50 vouchers of R200 in value will be awarded to randomly chosen survey participants. KLA will manage the competition and their decisions of pertaining to winners will be final.

Bank X will only collect and process the personal information in line with the purpose for which you provided it to us, or to enable it to comply with the requirements of specific laws that it is governed by. Bank X may also process your personal information in order to protect your or its legitimate interests. Bank X will not collect and process personal information about you that it does not need for this purpose. The general purpose for which it collects and processes your personal information includes but is not limited to usability testing which is a form of qualitative research.

If you have any questions or concerns, please do not hesitate to contact any of the following managers involved in this project.

APPENDIX E: Consistency table

Hyp	Hypothesis	Data collection method	Data analysis method
1	Chatbot failure with no recovery leads to a decrease in customer outcomes	Experimental app Post- survey	ANOVA; t-test
2a	Positive perceptions of recovery strategies *improve* brand intent to brand loyalty.	Experimental app Pre-survey Post- survey	ANOVA; t-test
2b	Positive perceptions of recovery strategies *improve* brand intent to re-use.	Experimental app Pre-survey Post- survey	ANOVA; t-test
2c	Positive perceptions of recovery strategies *improve* brand intent to satisfaction.	Experimental app Post- survey	ANOVA; t-test
3a	Handover to a human agent *leads to* a higher increase in *loyalty* over pre-failure levels than other recovery strategies.	Experimental app Pre-survey Post- survey	ANOVA; t-test
3b	A sense of distributive justice or equity rebalancing partly explains the change in loyalty following successful recovery strategies	Experimental app Pre-survey Post- survey	ANOVA; t-test
3c	A perception of interactional justice interacts with the relationship between the availability of a human online agent handover and improved *loyalty*.	Experimental app Pre survey Post- survey	ANOVA; t-test
3d	A perception of procedural justice *interacts with* the relationship between the service e recovery type and improved loyalty.	Experimental app Pre survey Post- survey	ANOVA; t-test
3e	Digitally native customers are more tolerant of alternative non-human digital interfaces as an initial recovery strategy rather than going straight to human agent, than would non digital natives.	Experimental app Pre survey Post- survey	ANOVA; t-test
3f	Customers *more tolerant* of service failure when asking the chatbot complex questions, than they would when asking the chatbot fundamental financial services questions.		

APPENDIX F: Ethics clearance certificate

Graduate School of Business Administration
University of the Witwatersrand, Johannesburg

Wits Business School

Wits Business School Ethics Committee
Constituted under the University Human Research Ethics Committee (Non-Medical)

Ethics Clearance Certificate

Ethics protocol number: WBS/DB9406170P/574

This certificate is only valid with a legitimate ethics protocol number and signed by the Researcher (below).

Project title	Chatbot says "Sorry I don't understand":Recovery strategies for chatbot service failure
Investigator / Researcher	Ms Claudette Greaves
Nature of Project	MM (Digital Business)
Decision of the Committee	Approved, provided stakeholders and participants are guaranteed confidentialty.
Issue Date of Certificate	2021-09-03
Expiry date	Date of submission of the project report
Chairperson	Prof Anthony Stacey ☎ +27 11 717 3587 +27 82 880 4531 anthony.stacey@wits.ac.za

Declaration by Researcher

One copy must be signed by the Researcher and returned to the Chairperson of the Wits Business School Ethics Committee.

I fully understand the conditions under which I am authorized to carry out the abovementioned research and I guarantee to ensure compliance with these conditions. Should any departure to be contemplated from the research procedure as approved I undertake to resubmit the protocol to the Committee.

Signature

27 FEBRUARY 2022
Date: