Universiteit Leiden

ICT in Business and the Public Sector

The Application of Behavioural Economics On In-App Surveys And Its Effect On Survey Response Rate: An Empirical Study

Name: Kyra Koper
Student-no: 2767902

Date: 24/05/2023

1st supervisor: Dr. Christoph Stettina
2nd supervisor: Dr. Yingjie Fan

MASTER'S THESIS

Leiden Institute of Advanced Computer Science (LIACS)
Leiden University
Niels Bohrweg 1
2333 CA Leiden
The Netherlands

Acknowledgements

First of all, I would like to express my deepest gratitude to my university supervisors Christoph Stettina and Yingjie Fan for their support during my master's thesis. Without their guidance, this research would not have been possible. Furthermore, words cannot express my gratitude to the team at Health Corp, who generously offered their knowledge, expertise and time to get the experiment running.

I am also grateful to my classmates for sharing their experiences and tips on writing a thesis and for their moral support throughout the process. Thanks should also go to the anonymous participants of the survey in this research.

Lastly, I'd like to mention and thank my family and best friend. Their constant support allowed me to stay motivated and continue my work.

Abstract

As organisations aim to stay ahead of competitors, they must keep their customers satisfied. Surveys allow organisations to receive insights into how customers experience their services and products. However, to gather as much feedback from customers as possible, it is important to receive a high survey response rate which in this day and age follows a decreasing trend.

To improve the survey response rate, this research experiments with behavioural economics. This is a concept that is usually used in user experience design to nudge customers into a desired behaviour, which is used to nudge customers to respond to surveys in this research. To do so, in this research behavioural economics has been explored and three behavioural economics principles have been selected to be tested on their influence on survey response rate. These three principles are authority bias, social influence, and loss aversion. A survey embedding these principles has been designed and implemented in an application of the healthcare organisation Health Corp. In addition to the three versions, a survey version that does not include any behavioural economics was implemented so that the performance of the other versions can be evaluated.

The results of the experiment are that the three behavioural economics principles received lower survey response rates than the survey without behavioural economics. The version without behavioural economics received the highest survey response rate, while the versions with behavioural economics received a much lower survey response rate. Conducting the chi-squared test indicated that the higher response rate did not happen by chance. As the meta-analysis resulted in high variability in results and no statistically significant evidence of behavioural economics being an effective method to boost the survey response rate, the recommendation is against the use of behavioural economics to boost the survey response rate and further research on this topic.

Table of Contents

1. Introduction

1.1. Background

Customers increasingly expect digital products and services to be more convenient (Bain, 2015). With the market competition simultaneously increasing, these customer expectations further encourage companies to satisfy the needs of their customers. For companies to create an understanding of these needs, more companies have started to invest in customer experience (CX) (Diebner, Malfara, Neher, Thompson, & Vancauwenberge, 2021). A 4% to 8% growth in revenue is the consequence of a company that excels in CX (Bain, 2015). Additionally, customer loyalty increases among existing customers and indirectly stimulates them to become promoters. The lifetime value of these promoting customers is 6 to 14 times larger than that of a customer on the opposite spectrum.

Customers' needs and expectations are dynamic. Thus, companies must pay special attention to these customer demands to meet them on time (Emdad & Bhatt, 2009). The flexibility of Information Technology (IT) allows for information, such as market changes, to be processed at a higher speed (Lucas & Olson, 1994). Using IT, companies have the opportunity to respond to these changes appropriately concerning their digital products and services. Moreover, IT enables the development of products at hyper speed (Cole, 2002). As a result, IT supports organizations in sustaining a competitive advantage in the market (Ness, 2008) (Byrd & Turner, 2000). The possession of a high level of IT infrastructure has been proven to have a positive effect on customer responsiveness (Emdad & Bhatt, 2009). Furthermore, IT allows for the enrichment of the customer journey as well as the CX (Hoyer, Kroschke, Schmitt, Kraume, & Shankar, 2020).

One of the techniques to measure how well a company is doing with CX is the survey research technique (De Keyser, Lemon, Klaus, & Keiningham, 2015) (Cano, Rusu, & Quiñones, 2020). Over the years, different survey modes have been used. In the early 1900s, the leading survey mode was the face-to-face interview (Wolf, Joye, Smith, Smith, & Fu, 2016). In a similar timeframe, postal surveys were also used for interviewers with a smaller budget. Following the postal surveys, the telephone survey found an increase in popularity in the 1980s and became the leading survey mode in that period. Nearing the 2000s and thus the rise of the Internet, web surveys (also known as Internet surveys) were introduced as a new data collection method. Additionally, with the increase in smartphone adoption, a new mode of survey method has been made possible – online mobile surveys (Biskirk & Andres, 2012). Online mobile surveys have several different approaches which are surveys taken through the smartphone's web browser, through an application or a combination of the two approaches.

Internet-based or online surveys are found to be especially advantageous for collecting data from respondents that are difficult to reach (Alessi & Martin, 2010). Other benefits of using this survey method are that it includes a shorter transmitting time and cuts down on delivery costs (Fan & Yan, 2010). Additionally, these surveys have a high level of scalability and allow for quick evaluation (Kohavi & Longbotham, 2017).

1.2. Issue

Amongst its many advantages, there are also multiple challenges. One major concern is the low response rates of an online survey which consequently threatens the quality of the survey result – a nonresponse bias (Sánchez-Fernández, Muñoz-Leiva, & Montoro-Ríos, 2012). In this situation, the risk is that the sample of responses may differ significantly from the nonrespondents, which leads to a biased perception of the sample population.

In the existing literature, studies have been conducted looking into the response rate of postal, email and telephone surveys, as well as web-based surveys. In these studies, several factors have already been identified that affect the response rate of surveys, such as the length of the survey, question types, the complexity of surveys, the survey design and the topic of the survey (O'Reilly-Shah, 2017; Fan & Yan, 2010). Also, many studies have been conducted comparing the response rate between these different survey methods. However, little scientific knowledge is available on how the usage of an in-app survey affects the response rate nor which factors affecting the response rate become significantly more important when an in-app survey is used. Thus, it is unclear how to optimize the response rate for in-app surveys.

With surveys used as a common technique to measure CX, behavioural scientists argue that CX heavily influences the perception that a customer has during touchpoints with an organization (Devine & Gilson, 2010). Behavioural economics (BE) is an umbrella term for several principles that may increase the positive interactions customers have with an organization, resulting in significantly improved customer satisfaction. Despite this, surprisingly few studies have been conducted to explore how BE may be applied in a survey setting. Lemon et al. (2016) mention that a deeper understanding is necessary on the role that BE plays in customer behaviour at certain touchpoints. This discussion leads to further curiosity on the effect of BE on in-app survey response rates.

1.3. Solution

This thesis focuses on the identification of BE principles that affect survey response rates regarding app-administered surveys. First, the concept of BE will be further explored. Then, principles will be identified within the BE concept that may influence the survey response rate. The thesis aims to offer insight into how the response rate of in-app surveys may be improved by providing an outcome of different BE principles that influence the survey response rate and whether they are recommended for future usage.

1.4. Research Aims and Objectives

The purpose of this thesis is to look at how different BE principles would affect the survey response rate. The concept of BE will be explored and introduced to collect information on how these principles influence the survey response rate. This thesis follows an action research methodology with its main aim to provide knowledge about how appropriate BE is to improve the survey response rate. This leads to the following explanatory research question:

What are the benefits and the challenges of applying behavioural economics in order to understand and improve the response rate of in-app surveys?

1.5. Scope

The healthcare company where this research was done will remain anonymous throughout this thesis and will hereby be referred to as Health Corp. Health Corp its existing problem has been identified regarding the response rate of their piloted survey. As a result, this research will be carried out in the form of a case study. The duration of this internship was 9 months. The research has a strict deadline as after the research, the researcher must continue for another semester to graduate.

Health Corp has developed an application for nurses which is the main focus of this case study. Several touchpoints of the survey have already been identified by Health Corp and will be used to trigger the survey. These touchpoints are the following:

- Prescription process;
- Patient registration process;
- Treatment order process.

The main topics that will be discussed in this thesis are as follows:

- Surveys;
- Customer Experience;
- Online Controlled Experimentation;
- A/B Testing;
- Behavioural Economics;
- Survey Response Rate.

1.6. Structure of the Thesis

The relevant literature on the topics explored in this thesis is presented in Chapter 2. In Chapter 3, the methodology utilized in this thesis is defined, and the execution and results of the research are discussed in Chapter 4. The discussion and limitations of the research are also touched upon in Chapter 4. Finally, Chapter 5 provides insight into the conclusion of the research.

2. Theoretical framework

This chapter aims to examine the literature relevant to the main topics of this thesis. Literature on the topics of customer experience (CX), behavioural economics (BE) and the survey response rate will be analysed and compounded.

2.1. Customer Experience

2.1.1. Definition

Customer Experience (CX) has been studied extensively for over half a century and through the years the concept has been given different definitions (De Keyser, Lemon, Klaus, & Keiningham, 2015).

The general concept of CX was first explored in 1956 by Lawrence Abbott. Abbott acknowledged that there is not a singular want for a product, but a composition of different wants (Abbott, 1956). Abbott explains that the product alone is not satisfactory enough for a customer, as more is expected than just the basic needs that come with the product. However, it is important to note that Abbott did not yet originate the term 'Customer Experience'.

The term CX itself was pioneered in 1994 by Carbone and Haeckel in the article 'Engineering Customer Experiences'. In this article, CX is defined as a consumer their impression of the products, services, or organisation (Carbone & Haeckel, 1994). The authors further build on this definition by explaining that humans filter perceived information into impressions. These impressions can be divided into rational impressions as well as emotional impressions. Additionally, the impressions can range from subliminal to apparent.

To explain CX, Pine and Gilmore (1998) explain the differences between commodities, goods, services and experiences. Commodities and goods are tangible, services are intangible, and experiences are described as memorable. While goods and services are externally perceived by the buyer, experiences are internal as each experience is unique to a buyer and influences a buyer emotionally, physically, intellectually, and sometimes even spiritually. The authors highlight that this implies that an experience can not be identical between different people. The economic distinction between the four economic offerings can be found in Figure 1.

Economic Distinctions				
Economic Offering	**Commodities**	**Goods**	**Services**	**Experiences**
Economy	Agrarian	Industrial	Service	Experience
Economic Function	Extract	Make	Deliver	Stage
Nature of Offering	Fungible	Tangible	Intangible	Memorable
Key Attribute	Natural	Standardized	Customized	Personal
Method of Supply	Stored in bulk	Inventoried after production	Delivered on demand	Revealed over a duration
Seller	Trader	Manufacturer	Provider	Stager
Buyer	Market	User	Client	Guest
Factors of Demand	Characteristics	Features	Benefits	Sensations

Figure 1: Economic Distinctions of the four economic offerings. Source: Pine & Gilmore (1998).

Shaw and Ivans (2002) derived from multiple interviews with senior business leaders that CX is composed of physical and emotional elements. The authors describe CX as a combination of the performance of an organisation and the emotions that are triggered by this. This combination is then set up against the customer's expectations across all touchpoints. The emphasis is put on the combination of the physical and the emotional, not distinctively one or the other. Additionally, CX is not only when a customer is inside of a shop, but any moment a customer interacts with an organization or its brand – the touchpoints.

Gentile et al. (2007) derived a definition of CX by combining that of other sources that are much in agreement with the previously named definition in this thesis. Here, CX is described as all the interactions that a customer has with a company and their product, consequently creating a reaction. The experience that this customer has is personal and this experience can be divided into different levels, which are rational, emotional, sensorial, physical, and spiritual. The way the customer experiences these different levels depends on the expectations that were set by the customer before their encounter with the company and their product.

Schwager and Meyer (2007) define CX similarly to Pine and Gilmore. The authors mention CX is a customer's contact with a company followed by an internal and personal response to this. This contact can be direct, which means purchasing a product or using a service. This contact can also be indirect which can take the form of reading articles in newspapers or advertisements on television.

Building on the definition of Gentile et al. (2007) and that of Schwager and Meyer (2007), Verhoef et al. (2009) explain that CX falls into a holistic category and categorized the different levels as described by Gentile et al. Instead, these levels are described as the cognitive, affective, emotional, social and physical levels that are part of the experience of a customer. These different levels all have elements that can be controlled by the retailed, yet also some that cannot. The authors explain that CX lingers with a consumer starting from the search for the product, all the way past the purchase and consumption of the product and lasting even after these phases.

According to De Keyser et al. (De Keyser, Lemon, Klaus, & Keiningham, 2015), CX is defined as being the combined term of cognitive, emotional, physical, sensorial and social elements that are evoked from the customer's interaction (both direct and indirect) with the representatives of an organisation.

Derived from existing knowledge about CX, Becker and Jaakkol (2020) defined 2 premises. The first is that CX is the unconscious and spontaneous response of a customer to all sale attempts by the organisation. The second is that the experience of a customer is ranged from ordinary to extraordinary and depends on how well the customer responds to these sale attempts.

Amongst the definitions that have been altered over time, several concepts are repeatedly used to define CX:

- Interactions between the customer and the organizations including its products, services and brands (Carbone & Haeckel, 1994) (Shaw & Ivens, 2002) (Gentile, Spiller,

& Noci, 2007) (Schwager & Meyer, 2007) (Verhoef, Roggeveen, Parasuraman, & Tsiros, 2009) (De Keyser, Lemon, Klaus, & Keiningham, 2015) (Becker & Jaakkola, 2020).

- The experience is personal and internal (Pine II & Gilmore, 1998) (Gentile, Spiller, & Noci, 2007) (Schwager & Meyer, 2007).
- The experience being compromised of *emotional, physical and intellectual* elements (Pine II & Gilmore, 1998) (Shaw & Ivens, 2002) (Gentile, Spiller, & Noci, 2007) (Verhoef, Roggeveen, Parasuraman, & Tsiros, 2009) (De Keyser, Lemon, Klaus, & Keiningham, 2015).

So, based on these most common agreements on the definition of CX a combined definition may be created, which is as follows:

Customer Experience (CX) is the personal and internal experience – compromised of emotional, physical and intellectual elements – of a customer during all interactions between the customer and an organization such as through their products, services, and brand.

2.1.2. Customer Experience (CX) Feedback

Corresponding to the used definition of CX in this thesis, Qualtrics (2022) states that CX is the combination of how a customer perceives, interacts with, and thinks about an organisation. Thoughts, interactions and perceptions are rather difficult to measure, so instead companies commonly use customer satisfaction to analyse how well the company can meet customer expectations. Customer satisfaction is suggested to be a mechanism supporting product quality, therefore improving the results that a business delivers (Anderson, Fornell, & Lehmann, 1994). With customer satisfaction supporting product quality, favourable business outcomes such as positive consumer buying intention and recommendations. As concluded by Anderson et al. (1997) there is a deep-rooted connection between customer satisfaction and desired business performance. This attractive consequence of customer satisfaction has generated interest from the business industry with the appeal of measuring customer satisfaction (Reichheld, 2006). At the current time, multiple models have been developed to attempt to measure CX.

2.1.2.1. (Swedish) Customer Satisfaction Barometer
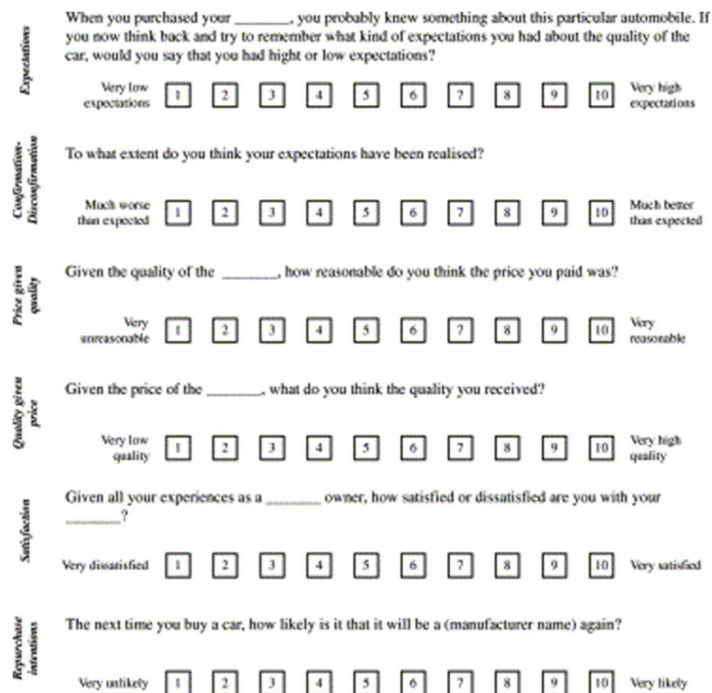
In 1992, a Swedish model developed by Claes Fornell was introduced: the Customer Satisfaction Barometer (CSB). This model was the first customer satisfaction measurement done on a national level rather than by individual companies (Fornell, 1992). Fornell described customer satisfaction as an alignment of the expectations of a customer before and after the purchase of a product. These expectations, in line with the Rational Expectations Theory, can predict a positive effect.

CSB is aimed to measure the quality of the output, which refers to the experience of a customer with the output. Fornell argues that customers that are facing costs that are rapidly changing should have a lower customer satisfaction than in industries where this is the opposite. As a first of its kind, CSB aims to measure three out of thirty previously established measures. These three measures are:

1. General satisfaction;
2. The confirmation of expectations, and;
3. The distance from the customer's hypothetical ideal product.

According to Fornell, CSB has the advantage, as opposed to traditional approaches, to measure the causes of satisfaction directly rather than in a confounded format. Additionally, the possibility of errors in the measurement of customer satisfaction is considered and the indicators that are used to measure customer satisfaction can be measured in such a way that the highest impact on loyalty and customer retention can be achieved (Fornell, 1992).

The first usage of the CBS was by Fornell to prove its usability was done through phone interviews. Fornell randomly selected customers of the largest companies in an industry and these customers were asked about one company only, except for the food and television broadcasting industries. During these eight-minute telephone interviews, Fornell would use a questionnaire with a 10-scale format. See Figure 2 for an example of the questionnaire. The survey response rate for this experiment was incredibly high at 95% (Fornell, 1992).

Figure 2: An example of the questionnaire for a customer of the auto industry.
Source: (Anderson, Fornell, & Lehmann, 1994).

2.1.2.2.   American Customer Satisfaction Index

With the changing economy, Fornell developed the CSB further in 1996 and called it the American Customer Satisfaction Index (ACSI). The ACSI evaluates the performance of companies on the quality of products and services as how they are experienced by the customer (Fornell, Johnson, Anders, Cha, & Bryant, 1996), much similar to that of the CSB. Figure 3 shows the ACSI model which shows the interrelationships between the different antecedents and consequences of the ASCI.
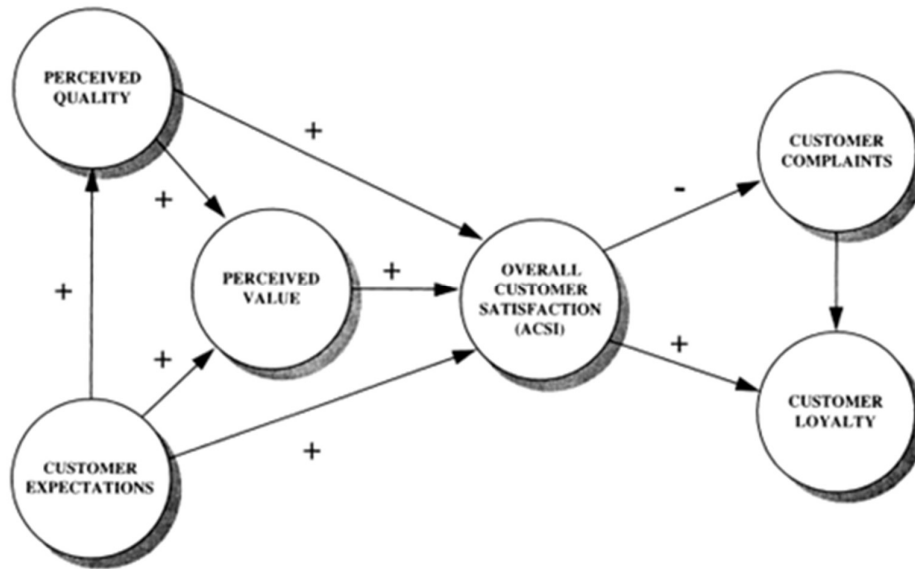
Figure 3: The American Customer Satisfaction Index (ACSI) Model. Source: (Fornell, Johnson, Anders, Cha, & Bryant, 1996).

Perceived quality (also embedding performance) is the evaluation of the most recent consumption experience of the market, which has a positive effect on both the perceived value and overall customer satisfaction.

The second antecedent is perceived value, referring to the relation between the price of the product and the quality of the product. This measure embeds the price information into the model, which leads to ease of comparing the results across different industries, as well as different companies within an industry.

The final antecedent is the targeted-customer expectations. This variable consists of both the pre-purchase and post-purchase expectations of the customer concerning the company's service or product. This antecedent captures the continuous relationship between a company and a consumer, making it a critical determinant of overall customer satisfaction.

Customer satisfaction also leads to consequences, of which one is a decrease in customer complaints (Fornell, Johnson, Anders, Cha, & Bryant, 1996). According to the exit-voice theory, the increase in customer satisfaction directly leads to a decrease in customer complaints (Hirschman, 1970). Additionally, the exit-voice theory explains that an increase in customer satisfaction also increases customer loyalty.

One study used the ACSI data released yearly by hotels, restaurants and airlines to examine the influence of customer satisfaction on a firm's financial performance (Sun & Kim, 2013). In this study, the ACSI represented the independent variable, profitability represented the dependent variable, and customer satisfaction represented the control variable. The study concluded that ACSI can influence the profit margin, return on assets, return on equity, and also the market-added value in a positive sense. Therefore, the study supports that customer satisfaction has a large impact on an organisation its profitability and value (Sun & Kim, 2013).

However, most studies created a questionnaire or survey themselves to gather data. These types of surveys differ from their administration. One study used telephone surveys to examine the consequences of citizen satisfaction (Van Ryzin, Muzzio, Immerwahr, Gulick, & Martinez, 2004). Another study used survey forms that were distributed among undergraduate students to analyse the relationship between customer expectations, satisfaction, complaints & loyalty, and perceived quality & value (Awwad, 2012). A third study used a structured questionnaire which was administered through personal in-home interviews to determine the usability of the ACSI in South Africa (Terblanche, 2006).

Across the different studies, a wide variety of survey response rates can be found which may be determined by its means of administration or other external factors. As such, an example of a response rate by an online survey was found to range between 3-7% (Wood, Siegel, & Feldman, 2008). Another online survey received a response rate of 17.2% (Nazi, 2010). In-person administered surveys had a significantly higher response rate of 75% (Awwad, 2012). Another study which followed an in-person administered method received a usable response rate of 63% (Deng, Yeh, & Sung, 2013).

2.1.2.3. European Customer Satisfaction Index

Eklöf built further onto the ASCI model, developing the European Customer Satisfaction Index (ECSI). In comparison to the ASCI model, the ECSI model has split the ASCI antecedent perceived quality into two types: perceived product quality, and perceived service quality (Eklöf, 2000). Additionally, the model introduced a new antecedent, which is the corporate image. As seen in Figure 4, the ESCI model only has one consequence which is customer loyalty.
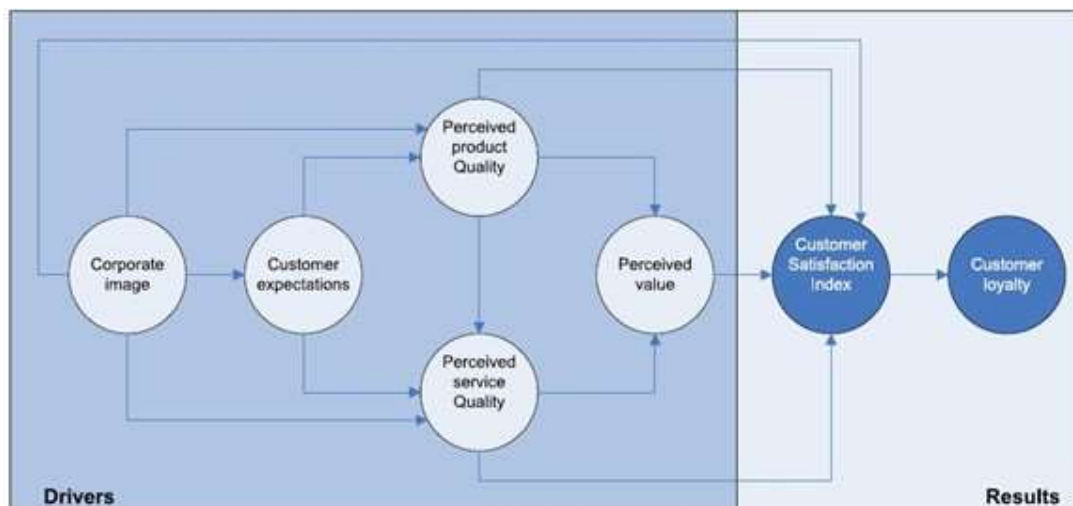
Figure 4: The ECSI model. Source: (van Haaften, 2017) .

Similar to the ASCI, most studies also administer surveys and questionnaires to gather data. Askaraizad & Babakhani (2015) used two different approaches to gather this data, which were in-person surveys and online surveys. Additionally, the authors used telephone recruitment to obtain agreement from the respondents as the responses were deemed anonymous. Although the authors did not enclose the response rate of their surveys, in-person surveys received exactly four times the amount of responses as compared to online surveys with 72

to 18. Another study used a postal administration method to gather data and received a survey response rate of 28%, which is assumed to be a good response rate for this type of survey administration method (O'Loughlin & Coenders, 2002)

2.1.2.4.   Net Promoter Score

Currently, one of the three most popular customer satisfaction metrics is the Net Promoter Score (NPS). Frederick Reichheld introduced the concept of the NPS in 2003 (Reichheld, 2003). In this article, Reichheld emphasizes the importance of customer loyalty in the growth of a company. In the context of surveys, one question stands out as the most important to measure, which is the question whether or not a customer would recommend the company to someone else. Net promoters are defined to be the amount of customers, in percentages, who are voluntarily promoting a brand or a company minus the percentage of customers who are 'detractors'. NPS is calculated with the results of a survey with the question mentioned previously. The response scale ranges from zero to ten, where zero implies the least likely scenario and where ten implies the most likely scenario that a consumer would recommend the company. Then, the promoters are calculated by determining the percentage of customers who responded with a nine or ten. The detractors are determined by calculating the percentage of customers who responded with zero through six. The percentage of detractors is then subtracted from the percentage of promoters which results in the NPS. A positive NPS implies company growth, while a negative NPS implies unhappy customers and the necessity to improve the NPS.

In 2018, Bain conducted an NPS benchmark study in the United States in multiple different industries (Bain, 2018). Figure 5 shows one of the cross-industry NPS benchmark results. However, it is important to note that while referencing benchmarks is relevant, factors such as country should be considered as they may impact the benchmarks.

Figure 5: NPS benchmarks for US life, property and casualty, and health insurers. Source: (Bain, 2018).

For this type of index, surveys are strictly the only method of data collection. However, as mentioned previously, the questions are different as opposed to that of the other approaches.

This is because the goal that NPS tries to achieve, which is to see if customers were to recommend a product or service, requires the survey to be designed in such a way. An example of a study where NPS was used, used an online survey format to ask the customer how likely the participant was to recommend the university and the followed courses to a friend (Schmatz, Wolf, & Landmann, 2015). This study using online survey administration received a response rate of 15.6%.

With the popularity of social media, social media channels such as Twitter provide an approach for approved businesses to create 'Customer Feedback Cards' to ask about how likely a customer is to recommend the business to a friend (Twitter, 2022). According to Twitter, these feedback requests have a survey response rate of 50% (Cuttica, 2016).

2.1.2.5. Customer Satisfaction Score

The Customer Satisfaction Score (CSAT) is also part of the three most popular customer satisfaction metrics. CSAT is best operationalized by monitoring the relationship between the expectation of the customers and the actual outcome (Buttle, 1994). This means that customer satisfaction may be predicted when the expectation is equal to that of the outcome. When the expectations are higher than the outcome, customer satisfaction is low. This also means that if the expectations are lower than the outcome, customer satisfaction is high. According to Qualtrics (2021), CSAT is measured through the use of surveys and requires specific question-and-answer formatting. The question is required to ask how satisfied the customer was with the good or service that they used or received. The answer must range between five answers:

1. Very Unsatisfied
2. Unsatisfied
3. Neutral
4. Satisfied
5. Very Satisfied.

Based on the answers received, a calculation is made to measure the CSAT. The number of satisfied customers, which are the customers that either responded with 'satisfied' or 'very satisfied', is divided by the total amount of survey responses that were received. The number is then multiplied by one hundred to receive a percentage. The benchmarks for the CSAT differ per industry. The American Customer Satisfaction Index website has created benchmarks across these different industries, such as retail, travel, healthcare, and finance (The ASCI, 2022).

Over time, creative ways to administer surveys have been found. Rather than the more common postal, telephone, interview or online surveys, a different take on giving feedback has been given by the firm Happy or Not. Happy or Not created a physical 'feedback terminal' so that customers can answer a survey question with the press of a button (BBC News, 2019).

See Figure 6 for an example. However, it is unclear what the response rate is like for this type of survey administration.

Figure 6: An example of the feedback terminal used at an airport. Source: (BBC News, 2019).

2.1.2.6. Customer Effort Score

In 2010 another customer satisfaction metric was introduced, the Customer Effort Score (CES), which is now one of the three most popular metrics. The authors of the CES argue that rather than receiving over-the-top service, customers instead want an easy and prompt solution to the problem that they are experiencing (Dixon, Freeman, & Toman, 2010). According to the authors, when comparing the predictive powers of CSAT, NPS, and CES, CES performs the best in the prediction for repurchasing as well as increased spending. To measure CES, the survey can ask the participant how much effort had to be applied to be able to succeed in their goal, like handling a request. Much like CSAT, the response scale follows a five-item scale, ranging from very low effort to very high effort.

However, the CES has received a lot of criticism for not accurately measuring customer satisfaction or customer retention. De Haan et al. (2015) researched the different measurements to compare their predictions for retention across different industries. The study concluded that CES was the worst-performing predictor for customer retention rate and does not seem to have a strong impact. Due to the clashing opinions on the effectiveness of the CES and its minor role in this thesis, it will not be discussed further.

2.1.3. Feedback Collection

Each company may have a different goal for tracking customer satisfaction or customer experience. Meyer & Schwager (2007) mention three different patterns that a company may choose to analyse. These different patterns are the 'Past', 'Present' and 'Potential' patterns. Each pattern differs in purpose, the frequency of the data collection, and the method used to collect data. For this thesis, the most relevant pattern type is the 'Present Pattern'-type, which is used to evaluate the current relationship between a customer and a company. This means that the company intends a long-term relationship with the customer. The data is usually

collected quarterly and the data is collected through web-based surveys with a moderated user base. Please see Appendix A – Pattern Types for Collecting Data, for more details on these pattern types.

An example of how customer feedback is relevant to a company is the case study of HiTouch. HiTouch had lost one of its largest customers to a competitor. They created a relationship survey to administer to their top customers to quickly find the touchpoints that their customers were disappointed in (Meyer & Schwager, 2007). As opposed to the initial positive opinion of the executive management at HiTouch, the company was informed about their defects and within a year time managed to reduce defections by an average of 16%.

2.2. Behavioural Economics

Behavioural Economics (BE) exists out of two concepts, 'behaviour' and 'economics'. Economics refers to the conceptualization that the economy is driven by the 'Homo Economicus' (Mullainathan & Thaler, 2000). In classical literature, the 'Homo Economicus' is described as an individual that only makes rational decisions and therefore cognitive and emotional influences do not influence the decision-making of a 'Homo Economicus'. This theory, which also goes under the name 'Rational Choice Theory', assumes that the Homo Economicus possesses perfect knowledge and maximizes self-interest utility (Schlüter, et al., 2017). However, experimental and empirical data provided evidence that this view was incorrect and out of this bloomed BE (Mullainathan & Thaler, 2000).

BE is a broad concept about the relation of psychology to economic decision-making (Kenton, 2020). According to Rudenko et al. (2020) BE is the study of actual behaviour and its underlying psychological elements that can influence an individual's decision-making process. BE studies the difference between the rational, or optimal, decision-making of individuals and the actual decisions made, and the consequence of this (Witynski, 2011). BE studies have shown that human beings do not strictly make rational decisions, even when they possess the right tools, information and knowledge that allows them to make rational decisions (Witynski, 2011). Individuals are influenced by emotions, environments and circumstances. These influences relate to an individual their cognitive ability and the psychological biases that are involved (Madrian, 2014).

In Nobel Prize winner Daniel Kahneman's New York bestseller book, the author discussed heuristics and aims to explain how minds work. According to Kahneman (2011), the mind consists of two systems. The first system is intuitive and unconscious and is the first system that is used when making decisions, as this requires the least amount of brain power possible. The second system makes choices based on logic and therefore is conscious. When the problem is too difficult for system 1 and requires logic, system 2 is used. The mind has developed shortcuts to make quick judgements, these are heuristics. As the mind is considered lazy, system 1 is preferably used. However, because of this, system 1 becomes overused and can therefore easily be tricked. Examples of heuristics named are the availability heuristic and the substitution heuristic. The availability heuristic is the overestimation of the probability something will occur because of repeated exposure to it. For example, when the newspaper reports money heists on multiple occasions, the assumption is made that money heists happen often even though the occurrence of a money heist is still quite rare. The

substitution heuristic is that instead of answering the question that was asked, an easier question similar to the original question is answered instead.

2.2.1. Behavioural Economics Framework

To select an appropriate framework for this thesis, existing frameworks were reviewed. Of these frameworks, the selection of one was based on criteria relating to literature. The MINDSPACE framework was the most used in literature. However, the other criteria can be seen in Table 1 and will each be explained further. For its scoring system, a framework can score between one and three based on its performance in each aspect, with one being the lowest score and three being the highest. So, as a final score, twelve is the highest score that can be achieved.

The 'Coverage of BE Principles'-aspect requires the framework to both have BE principles in the framework itself, as well as a broad coverage of these principles. Upon scoring a one for this aspect, the framework was immediately deemed unusable for this thesis.

The aspect 'Citations' refers to the number of citations received by other researchers. The score is given by benchmarking each framework against one another. For example, the MINDSPACE framework received around 850 citations on Google Scholar, while Deloitte published their framework on their website with an absence of citations. Therefore, the former scored three points, while the latter scored one point.

The 'Science-Based'-aspect refers to what degree the creators of the framework used science to justify their claims on the effectiveness of the framework.

Finally, the 'Application Example Available'-aspect scored the framework based on how well the framework explains how the principles within the framework should be applied.

Table 1: The criteria and scoring for selecting a BE framework.

Framework	Coverage of BE Principles	Citations	Science-Based	Application Example Available	Final score
MINDSPACE	3	3	3	3	**12**
EAST	2	3	3	3	**11**
NUDGES	2	3	3	3	**11**
BASIC	3	1	3	3	**10**
Learn, Test, Adapt	1	-	-	-	**1**
Define, Diagnose, Design, Test	1	-	-	-	**1**
Deloitte	3	1	3	2	**9**
Stepps (Contagious)	1	-	-	-	**1**
BER Framework	3	1	1	3	**8**

Table 1 shows multiple frameworks that conceptualise BE and that have been created and publicised by researchers. Following the scoring system, multiple frameworks were excluded immediately when their coverage of BE principles was too low. The highest-scoring frameworks are as follows:

- The MINDSPACE framework;
- The BASIC framework;
- The NUDGES framework, and;
- The EAST framework.

In the following paragraphs, the different frameworks will be discussed with a primary focus on the MINDSPACE framework (Dolan P. , et al., 2011) which scored the highest in the selection process and has been chosen for this thesis.

2.2.1.1.   The MINDSPACE Framework

Derived from BE is the MINDSPACE Framework, created by Dolan et al. (2011), a non-exhaustive mnemonic for describing nine BE principles that unconsciously influence an individual's behaviour and consequently their decision-making. As seen in Figure 7, MINDSPACE consists of:

1. Messenger;
2. Incentives;
3. Norms;
4. Defaults;
5. Salience;
6. Priming;
7. Affect;
8. Commitments, and;
9. Ego.

SALIENCE
Our attention is drawn to novel things that seem relevant to us.

DEFAULTS
We "go with the flow" of pre-set options.

PRIMING
Our actions are often influenced by subconscious cues.

NORMS
We are heavily influenced by what others do.

AFFECT
Our actions can be powerfully shaped by our emotional associations.

INCENTIVES
Our responses to incentives are shaped by predictable mental shortcuts, such as the strong desire to avoid losses.

COMMITMENTS
We seek to be consistent with our public promises and to reciprocate actions.

MESSENGER
We are heavily influenced by who is communicating information.

EGO
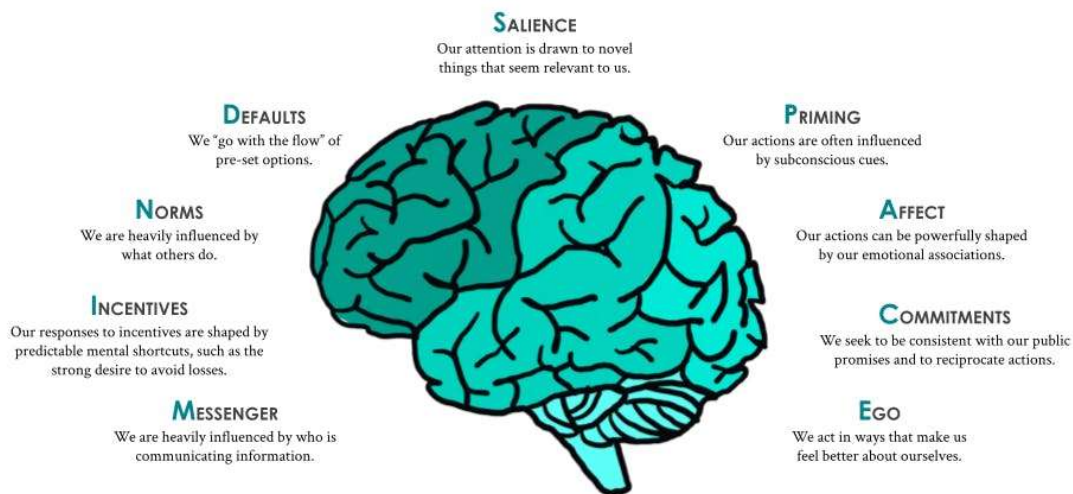We act in ways that make us feel better about ourselves.

Figure 7: The MINDSPACE Framework. Source: (The Decision Lab, 2018).

The **messenger** element in this framework, also more commonly known as the messenger effect, explains how the receiver of information weights the information based on the messenger or sender of that information. Studies have shown that people respond better to

information that is given by an expert (Dolan P. , Hallsworth, Halpern, King, & Vlaev, 2010). Studies surrounding the messenger effect can be found in multiple fields. Such as a study where it is shown that nurses comply with their authority figures, such as physicians, even if their authority figure is incorrect (Hofling, Brotzman, Dalrymple, Graves, & Pierce, 1966). Expertise, as well as peer effects, matter when it comes to the identity of the sender. A study has shown that among students, students are more likely to smoke when their peers smoke as opposed to their parents (Dolan P. , Hallsworth, Halpern, King, & Vlaev, 2010). Additionally, emotions affect how the received information is perceived. For example, when an expert delivers the information, but the receiver dislikes that expert, then the information will be disregarded. Receivers also use their rational and cognitive capabilities to decide how convincing the sender's information is which is heavily influenced by societal norms. Some principles related to messenger are 'formal authority' and 'halo effect'.

The **incentives** concept is directly related to the prospect theory. Incentives are used as a motivational cue for individuals to change their behaviour. According to Dolan et al. (2010) there exist 5 BE insights that relate to this topic. The first insight is that losses are perceived as worse than gains, which is the same as the statement in prospect theory: a potential gain has more value to an individual than a potential loss (Kahneman & Tversky, 1979). This is also called loss aversion. The second insight is that the reference point used is relevant. In other words, how an incentive is valued depends on the reference point. For example, when thinking about monetary incentives the amount may be compared to how much someone intended to earn in that week and how close the incentive is to that amount. It then matters how much an individual earns or loses compared to that reference point (Laibson & List, 2015). The third insight is that more weight is placed on small probabilities, even if the difference between probabilities is the same, e.g. 10-15% and 70-75% (Dolan P. , Hallsworth, Halpern, King, & Vlaev, 2010). The fourth insight is that the amount of money that is spent depends on the source of the money, such as salaries or savings which are called mental budgets or mental accounts. The final BE insight is that people prefer payments that are smaller and where the payoffs are almost immediately compared to payments that are large and the payoffs are in the future.

The third element of the MINDSPACE framework is **norms**. Norms refer to what is accepted in society or within a certain culture (Dolan P. , Hallsworth, Halpern, King, & Vlaev, 2010). These norms can be both conscious or unconscious and explicit or inexplicit. These kinds of norms are flexible as they depend on the behaviour of others around someone, therefore they can evolve, change and spread quickly.

The fourth element is **defaults**. This topic is described as options that are already selected when an individual does not make a choice themselves (Dolan P. , Hallsworth, Halpern, King, & Vlaev, 2010). An example of defaults is that in the Netherlands you are automatically registered as a donor unless you make the active choice of deregistering. However, BE explains that the mind is lazy and therefore will opt for the default choices regardless of the consequence (Kahneman, 2011). This can be seen in Figure 8. In 2020 the 'Donorwet' was implemented which is the law in the Netherlands where anyone over 18 years old by default becomes a donor (Donorregister, 2020). The figure shows the increase in the amount of

donors in the year 2020 compared to the previous years and that over the next two years, that number has been increasing. The blue line represents the amount of donors that actively opted in to become a donor, while the dark purple line represents the number of donors that were opted in automatically.
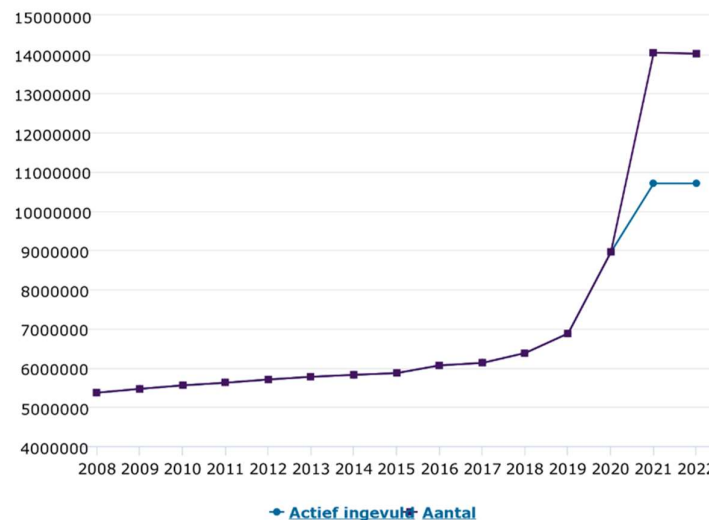

Figure 8: Amount of registered donors per year in the Netherlands. Source: (Donorregister, 2022).

Salience in the framework refers to the attention of people shifting to both novel and relevant things (Dolan P. , Hallsworth, Halpern, King, & Vlaev, 2010). The psychologists Taylor and Thompson (1982) describe salience as the occurrence that people give more weight to topics that they give more attention to which consequently affects their judgement. In the digital age, people are bombarded with data, which requires a lot of filtering as our attention can not be everywhere. Because of this, people instead focus on things that are novel and relevant, things that are well-understood or are relatable. For example, as given in the article by Dolan et al. (2010), the budget of the National Health Service becomes more 'salient' when it is expressed in something relatable or personal. So, instead of simply giving the whole budget amount, expressing it in amount per taxpayer makes the budget seem more salient. On many occasions, people look for an 'anchor' to base decisions on, much like a reference point.

Another element in MINDSPACE is **priming**. Cohn and Maréchal (2016) describe priming as the influence of the behaviour and judgement of an individual through subliminal messages which can be measured. Especially important is the keyword 'subtle'. There are many different kinds of primers, such as words, sights or smells (Dolan P. , Hallsworth, Halpern, King, & Vlaev, 2010). For example, in a study by Dijksterhuis and Bargh (2001) a group of people was exposed to words that related to old age. The consequence of this is that this group started to act accordingly to people of old age, such as walking slowly or experiencing forgetfulness.

The **affect** element relates to the emotional reaction of humans and is very potent in the decision-making process (Dolan P. , Hallsworth, Halpern, King, & Vlaev, 2010). Emotional responses to situations happen before an individual realizes what is it that they are reacting to. A common example is the mood that someone is in and how it influences how they respond

to something. If someone is already in a bad mood and someone bumps into them, they would respond to the person bumping into them more negatively than if the person was in a good mood. Emotions are so powerful that it actively influences an individual's perception. This means that when we see a thing, this thing is a beautiful thing or an ugly thing, and so on.

The second to last element of the framework is **commitment**. Commitment is what is necessary to be able to achieve long-term interests (Dolan P. , Hallsworth, Halpern, King, & Vlaev, 2010). Often to keep committed to a (long-term) goal, people think of a way to punish themselves if they do not commit to reaching the goal. James Clear (2018) describes in his book 'Atomic Habits' which steps should be followed to increase the commitment to goals. For example, Clear explains that one should not focus on the goal itself alone as the motivation will slip away once it is achieved. Rather, one should focus on the process to achieve the goal as this 'system' creates progress. People often fail to commit, because the progress is exponential and not linear. In Figure 9 the grey area between the expected progress and the actual progress marks the 'valley of disappointment'. According to Clear, this is the period when people stop committing to achieving the goal.

The Plateau of Latent Potential

Results

What actually happens

Valley of disappointment

What you think should happen

Time
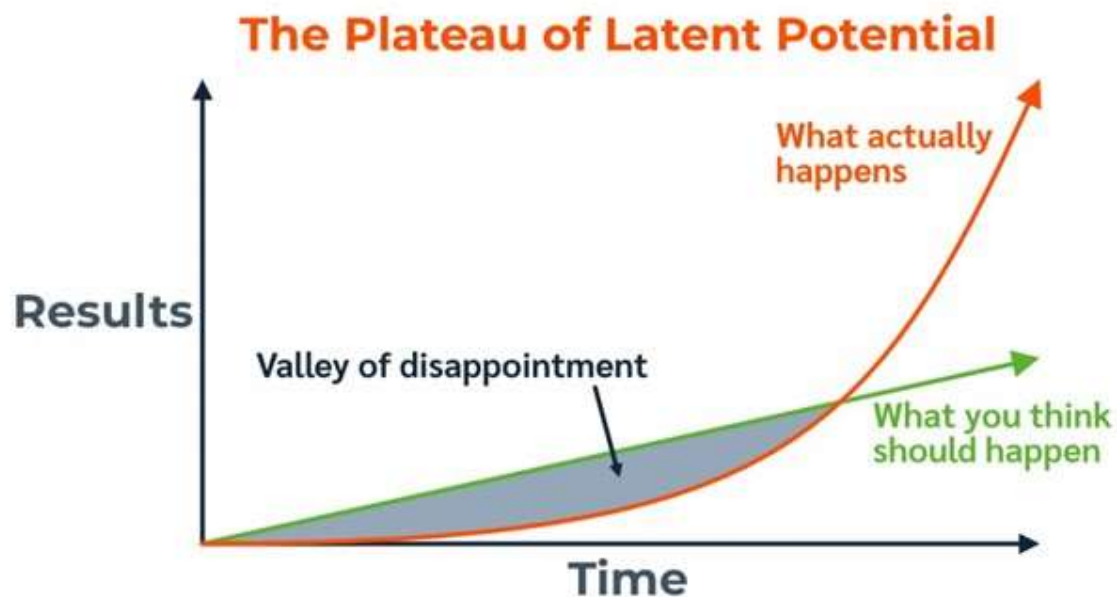
Figure 9: The Valley of Disappointment. Source: (Clear, 2018).

The final element of the framework is **ego**. People like to 'feed into their ego'. In other words, people behave in a certain way to make themselves feel good themselves and appear good to others (Dolan P. , Hallsworth, Halpern, King, & Vlaev, 2010). Additionally, when life is going smoothly, people praise themselves for it. When life is facing hardships, people blame others instead. The ego changes the way people see the world and a good example of this is fans and their memories of how their favourite team performed in a match. While teams are capable of the same fouls, to a fan the opposite team appears to have made more fouls than their favourite team.

The entire MINDSPACE framework, including its principles relating to each element, can be found in Figure 10.
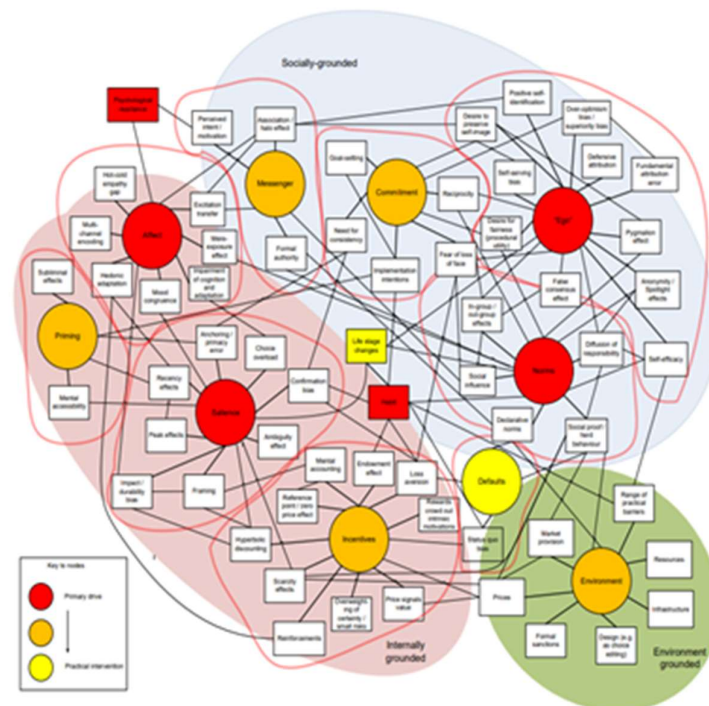
Figure 10: MINDSPACE mapped with related principles. Source: (Dolan P. , Hallsworth, Halpern, King, & Vlaev, 2010).

The definition of each principle embedded in the MINDSPACE framework can be found in Appendix B – MINDSPACE Principles Definitions.

2.2.1.2. The BASIC Framework

BASIC is a toolkit for policymakers and was created in 2019. This toolkit has multiple best practice tools, methods and guidelines to execute behavioural insight (BI) projects in its entirety and may be further supported by other BI frameworks (OECD, 2019). The framework consists of five stages as seen in Figure 11 below.

Figure 11: The BASIC-Framework (OECD, 2019).

The first phase, Behaviour, focuses on the problem definition that the policy wants to address. Furthermore, it helps with identifying and defining the behaviour aspect of that policy.

The analysis builds on the prior phase by identifying why people behave in certain ways. The BASIC framework refers to another framework, the Attention, Belief Formation, Choice, and Determination framework (ABCD). This framework further explains how each of the four aspects is linked to the Rational Choice theory.

Strategies is the third phase and support the identification of strategies to change the identified behaviours to that of the desired one. This phase highlights to make the next phase its interventions relevant enough to receive the attention of the individuals whose behaviour is aspired to be changed. Behaviour economics is at the centre of this phase.

In the Intervention phase, the strategies that were decided upon are tested to see which are most effective and therefore successful in reaching the desired policy. A common approach to test which intervention works the best is through the use of A/B Testing.

Finally, the Change phase uses the strategy that was found to perform the best in the previous phase to think about the long-term applications of the strategy. This is the critical phase where the direction of the BI project will be decided.

2.2.1.3.   The NUDGES Framework

Thaler & Sunstein (2009) published a book about their concept of 'nudges'. A nudge is defined as a method to get someone's attention. More specifically in this context, a nudge is a choice made in architectural design to change the behaviour of people in a predictable and desired way without making severe restrictions. The book discusses multiple BE principles, such as 'Anchoring', 'Defaults', 'Loss Aversion', and 'Status Quo Bias' among others. The book explains how these principles work and where and how they may be applied.

2.2.1.4.   The EAST Framework

In 2012, the EAST framework was introduced by the Behavioural Insights team. EAST is an acronym for 'Make It Easy', 'Make It Attractive', 'Make It Social', and 'Make It Timely' (BI Team, 2014). Each aspect has its own goal which requires different BE principles to sustain them.

Make It Easy explains that it is important to make options as easy as possible. It suggests the use of defaults, or a pre-set option, to increase the ease of making a decision. Additionally, it suggests to reduces the effort which makes increased uptake or survey response rates possible. Also, it explains to make the message that is tried to be conveyed as easy as possible too which often results in a large increase in survey response rate.

Make It Attractive argues that people are more willing to be moved to action when something draws their attention. Suggestions to draw attention are the use of images, colour or personalisation. Therefore, it is important to use an attractive design. Also, (financial) incentives are very effective to draw attention.

Thirdly, Make It Social explains the power of social contacts. Encouraging people to behave a certain way by showing how other people behave in certain situations is an example of how social contacts influence others. Additionally, encouraging people to make public commitments subtly forces people to follow and execute these commitments.

Finally, Make It Timely highlights the importance of timing for which the argument is that behaviour is easier to be altered when it is already disrupted by events. Also, it mentions how immediate changes in costs or benefits have a larger impact on people than costs or benefits that are delivered in the future. So, policymakers should analyse and see if it is possible to adjust those.

Furthermore, the framework follows a four-step method:

1. Define the outcome: what behaviour is to be influenced should be defined and how large this change should be.
2. Understand the context: during the situations where people show the behaviour in question, visit them and understand why people behave in such a way to design an intervention.
3. Build your intervention: use the framework to generate behavioural insights. This is likely to be iterative.
4. Test, learn, adapt: Test the interventions with treatment and control groups. Learn from the results and make changes in these interventions where necessary.

2.2.2. MINDSPACE Framework Principles Selection

The chosen framework for this thesis, the MINDSPACE framework, embodies a significant amount of BE principles. In Figure 10 in Paragraph 2.2.1.1, the entire MINDSPACE framework is visualized and shows the 52 BE principles. However, this thesis is unable to research all 52 principles within this framework, so just as with the framework selection, a selection has been made for all of the principles within this framework. The principles selection can be found in Appendix C – MINDSPACE Principle Selectio.

Each principle has been scored based on the degree of feasibility to test the principle, if the principle is logical in connection to the survey response rate topic, if the principle is logical in connection to the survey respondents' background, how well-grounded the principle is in literature, and how simple it is to test the principle.

A total of 5 principles scored full points on these aspects. These principles are Authority Bias, Social Proof, Social Influence, Framing, and Loss Aversion. However, Framing is so well-grounded in literature and taught about in academia when researching using surveys, that the decision has been made to exclude these principles from the final selection as it would not deliver much additional knowledge. Finally, Social Proof and Social Influence were found to be near-identical and so, the choice was made to use Social Influence among these two as the naming of this principle is more recognizable and familiar. In conclusion, the final selection of the principles is as follows:

- Authority Bias;
- Social Influence, and;
- Loss Aversion.

The following paragraphs will introduce these three BE principles further.

2.2.2.1. Authority Bias

Authority bias is a BE principle that may be defined as 'the increased likelihood of an individual being influenced by and agreeing with the opinion of another individual of status, without thinking further into the decision' (Eugenia Cadenas Sáez, 2020; Chakraborty, 2019; Tiwari, 2021).

In the late 1950s, authority bias had already been experimented with. In this experiment, led by Yale University professor Stale Milgram, authority figures asked the participants to perform

actions that went against their morals in order to study behaviour changes (Chakraborty, 2019). Interestingly, the majority of the participants followed each order by the authority figure despite clearly showing unwillingness to do so. Since this experiment, authority bias has received more attention. A more modern example of authority bias is the celebrity endorsement of products (Olenski, 2017; Chakraborty, 2019). The assumption that is made here is that because a celebrity is promoting a certain product, the product must be good. Consequently, people are influenced and convinced to buy the product.

However, despite the ability to influence behaviour, authority bias is also known to have a negative side to it. This negative aspect is people becoming increasingly lazy due to them accepting the opinions of authority figures without rethinking them (Chakraborty, 2019). For example, investors blindly accepting these opinions can cause them to be misled and invest in specific stocks, resulting in stocks moving in a specific manner (Tiwari, 2021). Furthermore, authority figures are often against creativity as this creates uncertainty (Gregoire & Kaufman, 2016). As people are influenced by these authority figures, this attitude creates a creative block and leads to a lack of innovation.

So, although authority bias can be used as an influential tool to change behaviour, it must also not be used excessively to avoid making large mistakes, or obstructing progress due to a lack of creativity.

2.2.2.2.   Social Influence

Much like authority bias, the social influence principle is also the influence of others on an individual. However, 'others' are typically referred to as another group (Deutsch & Gerard, 1955). More specifically, this group is increasingly more effective when the individual and the group have a good personal relationship. In an experiment conducted in 2012 that had 61 million participants, messages were sent out to individuals to attempt to influence their voting behaviour (Bond, 2012). Despite these messages being effective, the results of the experiment showed that social interactions about political opinions between close friends and those who received the message were able to influence the voting results to an even greater extent.

The influence of another group has multiple reasons, such as assuring the coherence of the group through good relations and appearing favourable to this group to receive a reward that the group can provide the individual. Cialdini (1998), explained these goals to be that of managing one's reputation and relationships. Often, this conformity of giving in to peers is due to negative emotions influencing this, such as feeling ashamed or a fear of being ridiculed and leading to deteriorating mental health (Lashbrook, 2000). A more common term for this negative aspect of social influence is called peer pressure.

Social influence is a powerful tool to influence the behaviour of others rapidly, however, peer pressure must be avoided as it may harm the mental health of an individual.

2.2.2.3.   Loss Aversion

In 1979, the term loss aversion was created by Amos Tversky and the Nobel-prize-winning psychologist Daniel Kahneman. This principle indicates that people prefer avoiding losses to gaining something (Kahneman & Tversky, 1979). This preference is apparent as an individual's

response to losing, for example, 5 dollars is 2 to 3 times bigger than it is compared to gaining 5 dollars.

The experiment that Kahneman and Tversky conducted was done by asking participants to choose between two different options. The first option was to be guaranteed to receive a certain amount of money. The second option was to have a probability to win twice as much money but with the off chance to gain nothing. From those participants, the majority preferred the first option, thus supporting that people prefer avoiding losses to possible gains.

On the contrary, loss aversion may also cause individuals to behave recklessly. In stocks, an individual may take a big risk in purchasing a certain stock in the hopes to avoid losses when the stocks will be sold. However, this may lead to the exact opposite which is significant losses (Kahneman & Tversky, 1992).

2.3. Survey Response Rate

When using questionnaires or surveys as a data collection method, researchers are heavily reliant on how willing people are to respond to the surveys (Baruch & Holtom, 2008). As defined by Frey (2018), the survey response rate is the ratio of how many people participated in the survey as compared to how many people were asked to participate in the survey.

The survey response rate is used to measure the quality of the data that is received. For example, non-respondents may differ severely from the respondents that did respond to the survey, this is called non-response bias. In a study of non-response bias, it was found that non-respondents of a questionnaire were less committed, had lower job satisfaction and viewed their relationship with their supervisor as bad compared to the respondents (Rogelberg, Luong, Sederburg, & Cristol, 2000). The difference between the non-respondents and respondents is so large, that the received data is found to be inaccurate and does not reflect the sample population.

In the literature, the response rate has been assigned different calculations. However, the most consensus is on a simple division where the amount of (usable) responses that are received are divided by the total amount of eligible individuals in a sample (Fincham, 2008) (Kviz, 1977). An issue that arises is the unfair penalization of individuals who are deemed to have 'unusable' responses (Harris, 2010). The main reason among many studies for deeming a response unusable is because of missing data. As an example, Nattinger et al. (2010) deemed any respondent with missing contact information as ineligible for the study. However, this does not apply to all questionnaires and there is no consensus on what is deemed 'ineligible' as this differs per study. The difference in the survey response rate between deeming respondents ineligible and eligible is substantial, as including the ineligible respondents in the study of Nattinger et al. would rise to a survey response rate of 70% as opposed to a survey response rate of 57%. Another study by Watanabe and Yamaguchi (1995) deemed unusable questionnaires to be those that missed data. 1150 questionnaires were distributed and 396 were returned, a response rate of 34.4%. However, after removing the unusable questionnaires, only 365 were usable responses, leading to a survey response rate of 31.7%.

It is common sense that a high survey response rate is deemed better than a low survey response rate, but for an extended time, it was unclear what the benchmark was for a 'good'

survey response rate. In 1999, Baruch researched the papers that reported survey response rate, which was not very common at the time. Baruch studied 175 different studies between 1975 and 1995 and concluded an average survey response rate of 55.6% (Baruch, 1999). Amongst these, the lowest reported survey response rate was 10% and the highest was 96%. Based on these results, Baruch suggested that in studies directed at top management, the survey response rate benchmark should be kept between 26-49%, whereas studies directed to mid-level management or the general population should follow a survey response rate benchmark of 40-80%. However, the assessed questionnaires here are strictly those of mail surveys or in-office distributed surveys. It is shown by Blumenberg & Barros (2018) that the use of web-based data collection resulted in a lower survey response rate compared to other methods such as mail, telephone or face-to-face survey administrations. A recent study, published in 2022 by Holtom et al., analysed the response rate of different survey methods between 2010 and 2020. This study also kept track of the amount of administered surveys per survey method. Between 2005 and 2020, the average reported to have increased from 48% to 68% (Holtom, Baruch, Aguinis, & Ballinger, 2022). In 2020, the highest response rate was found to be by in-person administered surveys with 73%. Following this, mail surveys received a response rate of 70%. Web-based surveys received a response rate of 65% and, surprisingly, phone surveys received the lowest response rate of 53%. However, it is also important to note that phone surveys are conducted very little each year, while online surveys substantially increased from 84 yearly in 2010 to 223 yearly in 2020.

Along with the heavy trend in web-based surveys, as suggested by the study of Holtom et al. (2022), web-based surveys are also highly relevant to this thesis. The reasoning is that the data-collection method for this thesis is a web-based survey. In other words, the data for this thesis will be collected through a survey that is designed and developed specifically to be distributed over the internet (Maymone, Venkatesh, Secemsky, Reddy, & Vashi, 2018). So, a heavier focus will be put on the survey response rate regarding this approach. A systematic review by Fan & Yan (2010) indicates that multiple factors affect the response rate of web surveys and contribute to their 'low' response rate. These factors are as follows:

- Survey development
    - The content of web questionnaires
    - The presentation of web questionnaires
- Survey delivery
    - Sampling methods
    - Contact delivery modes, such as reminders
    - Design of invitations
    - The usage of pre-notifications and reminders
    - Incentives
- Survey completion
    - Social-level factors, such as the degree of survey fatigue
    - Respondent-related factors, such as emotional stability
    - Design-related factors
- Survey return
    - Survey software

        o   Data safety.

Suggestions by Fan & Yan (Fan & Yan, 2010) are to first pilot the survey with a small number of respondents so that the results may be analysed and the questionnaire may be revised. Also, the authors suggest simplifying the access to the survey to reduce the effort of respondents to respond to the survey. Additionally, the authors highlight the importance of the background of the respondents. The technology illiteracy of the respondents should be considered as elementary school students may have difficulty in writing simple phrases and people in a higher age group may have difficulty using a computer. Finally, technical failures have a detrimental effect on response rate, so stable survey software should be used. In addition to the stability, the data safety of the software should be analysed, too.

3. Methodology

As first introduced in Chapter 1, the research question is as follows:

What are the benefits and the challenges of applying behavioural economics in order to understand and improve the response rate of in-app surveys?

To select an appropriate method to collect and analyse the data so that the research question may be answered, the Research Onion model by Saunders, Lewis & Thornhill (2019), as depicted in Figure 12, is used.
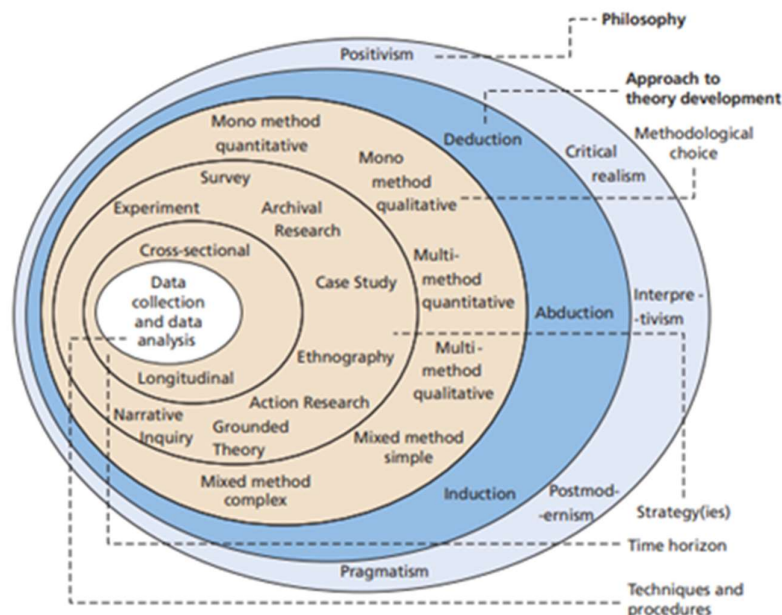
Figure 12: The research onion. Source: (Saunders, Lewis, & Thornhill, 2019).

Starting at the outer layer, this study follows the pragmatism philosophy. The pragmatism philosophy challenges theories in specific contexts to collect the data that is necessary to address the research problem and research question (Saunders, Lewis, & Thornhill, 2019). This research aims to assess the benefits and challenges surrounding the theory of BE in the

context of improving the in-app survey response rate. Therefore, an existing theory is tested using a deductive approach (Casula, Rangarajan, & Shields, 2021).

This research follows an action research methodology. Action research follows a systematic procedure which starts with the identification and definition of a problem and its context (Millis, 2017). Based on the problem, an intervention is developed and consequently implemented. The data from the implementation is collected and analysed and the results of this are reported. Kemmis, McTagger & Nixon (2013) describe this methodology as a 'spiral of action research', which highlights the iterative process of conducting research following an order of first planning the intervention, then observing the results after the implementation and finally reflecting on these results.

Despite its relevancy regarding results, action research also has been criticized for its lack of rigorousness. So, Davison et al. (2004) introduced the concept of canonical action research which proposes 5 principles. These principles function as a practical guide which they argue is missing in action research. So, these 5 principles will be followed to boost the rigorousness of this research. The first principle is the Principle of the Researcher-Client Agreement. This research is conducted in collaboration with Health Corp, a healthcare company. Therefore, a common understanding of the area of focus and approach to the research is important in addition to transparency about methods of data collection. The second principle is the Canonical Action Research Process Model as shown in Figure 13. Similar to the process of action research described by Millis (2017) and Kemmis et al. (2013), the process follows a 5 step order: diagnosis, planning, evaluation, and reflection (Davison, Martinsons, & Kock, 2004). This research follows each step in this process. The third step in this process is the Principle of Theory. This research is guided by the theory of BE and the intervention used is derived from this theory to evaluate the outcome. The fourth principle is the Principle of Change through Action which is to make certain interventions to change the current situation to a more desirable one. The research aims to fulfil this by implementing a BE solution to improve the response rate of surveys. Finally, the last principle is the Principle of Learning through Reflection. The stakeholders of Health Corp were regularly updated on the findings of the research.

For this study, quantitative data is necessary to achieve the aim to answer the research question. Quantitative data is required for an understanding of the impact that BE principles have on the survey response rate.
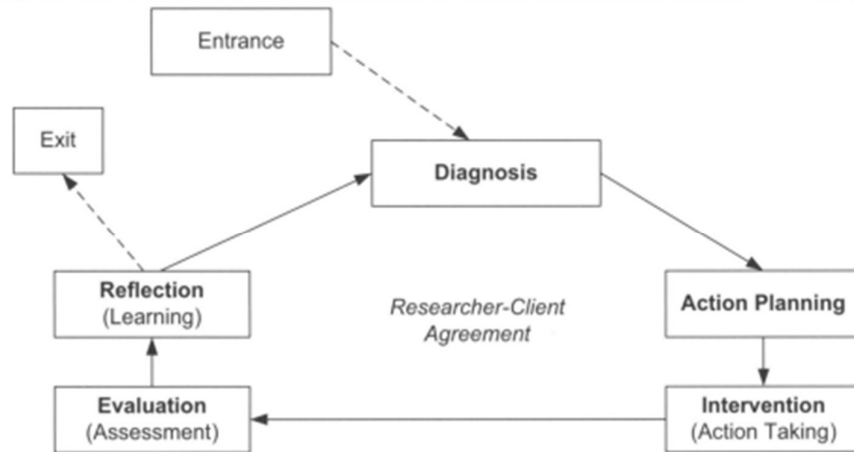
Figure 13: Canonical Action Research Process Model. Source: (Davison, Martinsons, & Ou, 2012).

3.1. Case Selection

This research was done in collaboration with Health Corp which will remain anonymous in this thesis as a result of a non-disclosure agreement between the researcher and Health Corp. Following a previous collaboration in 2021, a representative of Health Corp expressed their interest in the topic of BE as well as the lack of insight into customer satisfaction. After careful consideration of the current situation of Health Corp and its capabilities regarding available resources to conduct the research, the following topics were made clear: the focus of the research, research approach, Health Corp contacts and roles, the objectives, and the data collection and analysis methods.

Daily stand-up meetings that follow the Scrum methodology were attended daily to gather an understanding of the current situation before starting with defining an intervention. Additionally, one-on-one meetings were held with employees at Health Corp that were already working on the topic to gather more topic-specific information, such as information on BE and the current situation regarding attempts to gain insight into customer satisfaction.

Health Corp caters their products and services to the needs of their customers, which is especially important in this industry. This further emphasizes the importance of gaining these insights and attempting to improve the situation where possible.

3.2. Data Collection and Data Sources

To gather the quantitative data, this research used a combination of different data collection techniques. These techniques are online controlled experimentation in the context of A/B testing. The A/B testing was done by using different versions of surveys that embed the chosen BE principles as mentioned in Paragraph 2.2.2. of this thesis. These principles are seen as the treatment groups in the experiment, while a version of the survey without any BE principles was used as the control group. The choice of using A/B testing is that A/B testing allows as little performance variance as possible as each version of the survey ran separately, but simultaneously, over a selected but equally distributed sample size. The survey was active for multiple months, gathering data about the response rate of Health Corp's application users

(nurses) to the survey. So, in this research, the survey itself is not the data collection method, but a tool to conduct the experiment with. For the data analysis, the survey response rate of each principle was analysed and used in the evaluation and performance of each principle.

3.3. Research Design

As this research will be following the canonical action research process as defined by Davison et al. (2012), the research will be performed in five consecutive phases, as shown in Figure 13:

1. Diagnosis;
2. Action Planning;
3. Action Taking;
4. Evaluation, and;
5. Reflection.

Each phase will be briefly explained in the following paragraphs. Chapter 4 will go further into the discrete activities that were taken place during this research.

3.3.1. Diagnosis

In the first stage of the research, the current situation was diagnosed. This was done through conducting meetings with other employees who have been working on the CX topic. These meetings were held to gain more information about the previously brainstormed ideas to tackle the lack of insight into customer satisfaction, any steps that may have already been taken and that can be used for the research or to further work on, the general context of where to place the evaluation in Health Corp, current knowledge about BE within Health Corp, and what the resources of Health Corp were available and could be used (both in the sense of human resources as well as technology).

3.3.2. Action Planning

To gather more knowledge about which BE to use for the research, an intensive literature study was performed which was followed by a selection process of an appropriate BE framework to use for the study, which is further elaborated on in Paragraph 2.2.1. The principles introduced in this framework were chosen through another selection process based on certain criteria as mentioned in Paragraph 2.2.2. Multiple survey tools were evaluated and ultimately a survey created by Health Corp themselves was chosen. To prepare for the action-taking phase, multiple meetings with designers were held to have a design professionally created for each of the chosen BE principles, as well as one design for the control group. Additionally, meetings were held with developers to prepare the environment of the application for the implementation of the survey and discuss possibilities for the survey logic, such as distributing the survey among equal sample sizes, the logic of the survey appearing in the app, and the extraction of the data into an Excel-sheet.

A constraint that was found early in the research was the topic of privacy. Due to the health care industry and its high importance in privacy, multiple meetings were necessary to be conducted with the legal department of Health Corp to ensure that the data that was intended to be extracted was within the privacy policy stated by Health Corp. As this was not the case, the privacy policy had been updated.

Finally, when the survey design, the survey logic, the privacy policy, and the data extraction report were in place, the implementation of the survey into the application was added as a task into the DevOps system used by Health Corp.

3.3.3. Action Taking

As the survey had been designed and the privacy policy was adjusted, the survey versions were able to be implemented into the application of Health Corp. Each survey has been implemented at the same time to reduce the influence of external factors as much as possible. For the implementation process, the survey designs that were created were handed over to the developers along with the logic for the survey showing up to the users. After the implementation, the surveys went live on the 10th of October and the data collection started.

After running the surveys for a few months, a total of 94 responses were received across all versions of the survey. As a result, the control group received the highest survey response rate of 25%, and authority bias, social influence, and loss aversion received a survey response rate of 9%, 7%, and 11% respectively.

3.3.4. Evaluation

In the evaluation phase, the entire research was evaluated based on the benefits and challenges of the used experimentation method (A/B testing), as well as the limitations of the research. Following this, the areas of improvement such as population size and further research on other BE principles are discussed.

3.3.5. Reflection

Following the results of the data collection, the results were evaluated in this stage. The data was analysed to give potential explanations on why the versions with the BE principles received a lower survey response rate than the control group. In this analysis, the countries of the participants and the processes in which the surveys were shown were looked further into. Furthermore, external factors that could have influenced the performance of the surveys were discussed. Finally, this chapter also discusses what further work should be done on the topic of BE and its effectiveness on survey response rate.

4. Action Research

The major events of the research are listed in the following timeline:

- February 2022: Carrying out literature research & participating in meetings with employees for context and problem orientation
- February 24th, 2022: Collecting and analysing existing a BE-frameworks
- March 2022: Making the final selection of a BE-framework
- March 7th, 2022: Creating the data tracking sheet in Excel to collect data on the in-app survey response rate
- March 24th – 25th, 2022: Defining the in-app survey pop-up logic
- April 2022: Selecting the BE principles to use for the in-app survey designs
- April 9th, 2022: Creating a mock-up design of the in-app surveys
- May 2022: Discussing the privacy policy surrounding the in-app survey with the legal department
- June 2022: Handing over the design of the in-app survey to professional designers
- July 2022: Finalizing the design of the in-app survey
- August 2022: Implementing the in-app survey
- October 2022 – December 2022: Collecting the in-app survey response rate data
- January 2023 – March 2023: Evaluating the results, discussing findings and concluding the research

This research follows the canonical action research approach and followed the 5 phases: diagnosis, action planning, action taking, evaluating and reflection. The following paragraphs will discuss these topics further.

4.1. Diagnosis

In the Diagnosis section, the context of the research will be discussed, such as the problem Health Corp is facing, the reasoning as to why this is a problem, the current knowledge of Health Corp, and to what extent Health Corp has taken action to find a solution to this problem.

Context

In the initial phase of the research, the first steps were to create a starting foundation to begin the research. The most relevant aspects of Health Corp had to be explored such as their processes, communication points, the roles of the employees, their products, customers, tools and more. The supervisor of Health Corp created invitations to the most important meetings that their team participated in on Microsoft Teams, such as Daily Scrums, product demos at the end of a sprint, sprint reviews and retrospectives. This gave a clear introduction to the project management framework used by Health Corp, Scrum. These meetings gave an insight into the team their day-to-day activities and communication practices.

Product

The supervisor also explained the product that Health Corp provides, which is a home dialysis machine. This machine is created so that patients whose kidneys are not working normally can give themselves the treatment from home, rather than having to travel to hospitals or clinics to get the treatment done. The products that are used by the machine are also provided

by Health Corp and this is done in batches that need to be reordered by the patient through the mobile application. Missing a dialysis treatment leads to poor blood filtration which may then lead further lead to heart attacks or even death. So, it is extremely important for the patient to reorder the products and for Health Corp to deliver these in time.

Problem

Furthermore, the supervisor introduced multiple employees that were part of the team, the UX designers and developers, so that they would be able to introduce other aspects of Health Corp. During the meetings with them, the designers explained that they work on an application that supports the home dialysis product of Health Corp creates. This application has a mobile and a web version, where the mobile version is strictly available for the patients. Nurses have both the mobile and web version available to them. For the patients, the application supports them in creating orders for products that are necessary to continue to use the machine from home. They can see what products they need to order, what the deadline for ordering a product in time is and when the order will be delivered. However, as the nurses are the targeted audience for this research, the patients will not be asked to participate in the administered surveys of this research but will be mentioned throughout the research. The application has more functions for nurses and allows them to see the tasks that they need to do, and each task has a priority attached. They can manage their patients and see their orders and treatment status. But, Health Corp recently had been receiving complaints through their customer support about, for example, long loading times, the necessity for certain features, and missing fields in the application. Additionally, whenever they introduced new features for the application, they were unsure if this truly brought value to the customer. In short, they were receiving too little customer feedback to understand their customer satisfaction.

Collaboration

In addition to the application, they explained that the main tool that they use for sharing work is called Miro. This is a web-based whiteboard tool that allows teams to work together. Miro has a lot of features such as allowing the user to create notes, insert images and videos, and create project management-related designs such as sprints and Kanban boards. Figure 14 shows how the Miro board was used throughout the research. As Health Corp had identified that they did not have an insight into their customer satisfaction, the designers outlined the points of conversion where they deemed the feedback to be necessary and this was done in Miro. These points referred to a specific point in a process to gather feedback, such as right after a patient created an order to find out whether the patient was satisfied with this process. The technique that was decided to use to get this feedback was through the means of a survey. The designers reviewed how other organizations created surveys to receive feedback on customer satisfaction and thought of questions that would help them gather the information necessary. The overview of the identified points of conversion and the questions can be found in Appendix E – Identified Points of Conversion. Additionally, the designers and developers had thought out the logic that would control how often a survey would show up to the user of the application. The overview of the logic was created in Excel rather than Miro and it can be found in Appendix F – Logic Created by Health Corp.
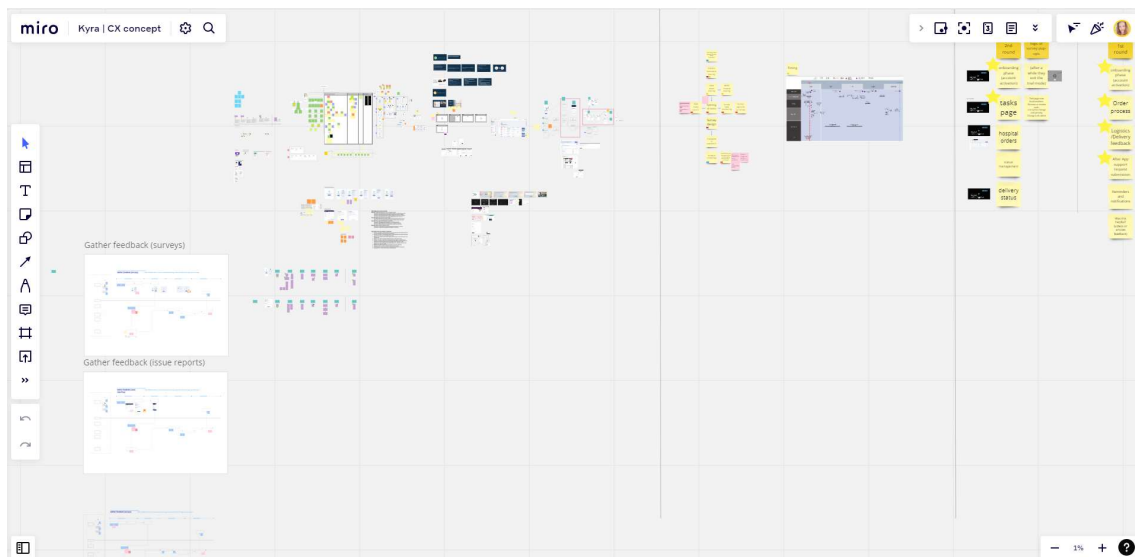
Figure 14: Visualization of how the Miro board was used throughout the research.

Research Outline

When this research started, no feedback mechanism was running yet. However, Health Corp wanted to stay ahead of their competitors regarding customer satisfaction so that it can retain its clients. This means that Health Corp had to know what could be improved and if the improvements are truly helping in solving potential issues. In addition to this, ensuring that the products and services provided are up to the customer's standards is increasingly more important when concerns health and well-being. Also, ensuring that the application can support a nurse in their daily tasks would boost their customer satisfaction, too. As mentioned, Health Corp had no insight into the potential issues or delights of the customer. Therefore, it has become a priority for Health Corp to implement a feedback mechanism, the survey, as quickly as possible. However, as mentioned in Paragraph 1.2, surveys that are conducted online suffer from a low response rate. With the high urgency of receiving feedback, it also became important to ensure a high survey response rate through certain means. Together with the designers, multiple brainstorming sessions were held on possible ways to improve the survey response rate. A senior designer with a high interest in BE was added to the sessions and introduced the topic further. Here, different principles that were used in user experience design were discussed and judged based on their potential for increasing the survey response rate. The result of this brainstorming session can be found in Figure 15. After going through the literature on the uses of BE, the final decision was made to use this as a means to improve the survey response rate.

Figure 15: Result of the brainstorming session about Behavioural Economics.

4.2. Action Planning

In the second phase of the canonical action research approach, the focus lies on the design of the survey to get this ready for implementation. This section will discuss the thought process behind the selection of the BE framework and the BE principles. The design of the survey is also touched upon in this paragraph. Additionally, it will also discuss the intervention that was necessary to protect the privacy rights of the surveyed individuals. Finally, the creation of the Excel document to store the data is explained in this section, too.

Behavioural Economics Framework Selection

In the previous phase, BE was explored. However, during the literature research on BE and brainstorming sessions with the designers at Health Corp, it came to light that the concept of BE has many underlying principles and best practices. The result of this brainstorming session can be seen in Appendix G – Brainstorming Sessions BE. To capture a broad, but set view of these principles, a framework had to be selected to gather these. During the literature research, multiple BE frameworks were found and evaluated based on criteria. This selection process can be found in Paragraph 2.2.1. The MINDSPACE framework was selected after carefully evaluating the frameworks based on their coverage of BE principles, the number of citations, the degree of how science-based the framework is, and the availability of application examples.

Still, the selection of the framework alone was not enough to prepare for the experiment. The chosen framework has 52 different principles which cannot all be used for the experiment. So,

following this, a selection had to be made of which principles would have the most potential in improving the survey response rate.

Behavioural Economics Principle Selection

After selecting the MINDSPACE framework, the principles part of the framework had to be selected to embed in the design of the surveys. As 52 principles were too many, a selection had to be made of at least 3 to 4 principles to keep the experiment effective, and also efficient. Just 1 or 2 principles would be inefficient as there are so many principles available, but only a few could be tested. It would also be ineffective to show the effect of BE on the survey response rate, as so few principles would not be able to represent the concept of BE well. However, too many principles would cause a lot of work to create designs for each principle, even for principles that may not be that appropriate for this experiment. A selection process that used specific criteria was followed and is discussed in detail in Paragraph 2.2.2. The final selection of the principles that received the highest scoring was as follows: authority bias, social influence, and loss aversion.

However, because it was uncertain if BE would truly improve the survey response rate, it became important to measure the performance of BE when applying this to a survey. So, choosing a method of measuring the performance was necessary.

A/B Testing

After selecting the BE principles to be used in the experiment, a method of measuring performance regarding survey response rate became crucial. After discussing the topic of measuring the performance of different principles with the designers, developers and supervisor, a quick decision was made to test the performance of BE through A/B testing. This way, the response rate of the survey with and without BE could be measured and compared through the means of different versions.

In literature, A/B testing is often referred to by its more formal term which is an online controlled experiment (Larsen, Stallrich, Sengupta, & Deng, 2022). However, in this thesis, the term A/B testing will be continued to be used. A/B testing is an experimentation method that is often used in the online marketing, application development and web design fields (Kohavi & Thomke, 2017). In this type of experimentation method, two or more versions of a product or web page are developed where one version is the control group, meaning that there is no change made to the design, while the other versions are the treatment group (Kohavi & Longbotham, 2017). The treatment group are the versions that differ in their design (colours, fonts, images), wording, or other factors that affect (consumer) behaviour, which may be how likely an individual is to purchase a product, or in this research how likely an individual is to respond to a survey. Even different versions of an application, such as mobile or web apps, can be used. So, here the assumption is that the different versions will result in different levels of performance after running for a set period. The outcome is compared and measured with different metrics such as conversion rates, click-through rates, engagement, customer satisfaction, or in this thesis survey response rate (Kohavi & Longbotham, 2017; Polonioli, et al., 2022).

However, A/B testing is not without its limitations. For example, it is difficult to control for external factors that may influence the outcome of the test, such as seasonality (Polonioli, et al., 2022; Fabijan, Dmitriev, Olsson, & Bosch, 2016). Furthermore, because the test is generally done with a limited sample of users, the results may not be generalizable to the general population. Despite its shortcomings, A/B testing has proven to be an effective tool for enhancing website design and marketing strategies. It offers valuable insights into how changes in design might affect user behaviour and performance. A/B testing is a key step in developing an effective online product or website. In addition, because A/B testing can point out benefits and challenges, its result will support the answering of the research question on what the benefits and challenges are of applying BE in the context of attempting to improve the survey response rate.

In this research, this methodology will be used to measure the performance of four different versions of the survey. In the control group, the survey version must be bland where no BE will be applied. Then, for the treatment groups, each version will be based on the previously selected BE principles. So, this means that there will be a version representing authority bias, social influence, and loss aversion. The survey response rate of each version in the treatment group will be compared to that of the control group to evaluate the performance of the BE principle in a survey setting. Naturally, if the survey response rate of a version in the treatment group is higher than that of the control group then the BE principle is effective. But, if it is the opposite, then the BE principle is ineffective.

Question Formatting

Before the researcher joined Health Corp, the designers had already investigated which processes they deemed necessary to receive feedback from. These processes were the password reset process, the prescription process, the patient registration process, and the order process. For each process, the designers had already created a question with an answering format that was inspired by other surveys. This iteration of these questions can be found in Figure 16 and Figure 17. The questions were different for most processes, such as asking to rate the nurse's most recent experience regarding the process, asking how easy the nurse found the process, or asking to what extent the nurse agrees with a certain statement. In addition to these questions, a field was left open to ask for the reason behind the given option.

However, the designers were unsure if this truly would be the best practice, so the researcher did more research on the best practices for formulating a good survey question. In discussion together with the supervisor and with the designers on what information would be most important to be gained from the questions in the survey, it was mentioned that it was specifically important to know how the nurses were experiencing the survey to know what they could improve. Multiple mock-ups were created to see which type of questioning would look best for this in combination with the best practices. Examples of these mock-ups can be found in Appendix H – Survey Questions Mock-Ups.

Figure 16: First part of the iteration of survey questions done by the employees as Health Corp.

Figure 17: Second part of the iteration of survey questions done by the employees as Health Corp.

Finally, it was decided that asking how easy the process was to the nurses would be the best option and was able to be used across all the different processes similarly. To get more information on what should be improved, an optional open field was added to ask the reason for the scoring. The reason that this is optional, is that it requires extra effort from the participants to fill it in. To avoid non-responses due to this, the field was created as an optional input. The questioning of the open field had gone through multiple revisions. The original question was phrased as '*What is the main reason for your score?*'. However, as this may not receive the answer that gives Health Corp enough input to make changes in the app, the

41

question was changed to '*What can be improved to change your rating by 1 point*?'. Still, as this question became too long, it was scrapped. So, it was changed to '*Any recommendations to improve the process?*'. This questioning was more straightforward and clear enough to be more successful in receiving a useful response.

Survey Tool

Consequently, a decision had to be made about which tool had to be used to implement the survey. As Health Corp had already been wanting to implement a survey, initially SurveyMonkey was chosen as the tool to use. However, due to its costs, LimeSurvey had been identified to be used instead. But multiple concerns came up during meetings about using this tool. The first concern was that the surveys were going to be sent out by e-mail. However, if customers are not likely to read their e-mails or if the e-mail containing the survey would end up in their spam folder, then they would not be responding to this which may influence the survey response rate of the experiment. Additionally, this would not be in line with the research which attempts to measure the effect of BE on in-app administered surveys. Then, instead of using e-mails, the survey would be implemented into the application itself, it would require much more effort on the developer side as this is a third-party tool. Finally, to be able to apply BE it was necessary to customize the survey to a certain degree. However, LimeSurvey lacked the customizability options in order to do so. Following a meeting with the developers, designers, and supervisor, where these concerns were taken into consideration, a decision was made to create a survey from the top. This meant that the survey had to be designed, and the logic and the survey had to be implemented by the developers.

In-App Survey Design

To design the survey, meetings were held with the designers at Health Corp to define the UX design. As the decision had been made to create a survey rather than using a third-party tool, more thought had to be put into the process. This meant that decisions had to be made on priorities. The first decision had to be made between creating a survey for the web or mobile, as the product by Health Corp was able to run on both. However, upon discussion with the supervisor and the designer, it was mentioned that the nurses barely use the application on their mobile. So, creating a design for the mobile version for nurses did not seem efficient. Therefore, the decision was made to make a design for the web version of the application instead.

As it became clear that a design had to be made that would be appropriate for the web version of the application, brainstorming sessions were held about which surveys by other companies could be used for the design. The outcome of such a brainstorming session can be found in Appendix I – Brainstorming Session Survey Design. Although inspiration was taken from these designs, no design was selected to be used. There were multiple ways to have the survey show up, such as slowly sliding up from the bottom, or it appearing in the middle of the screen. However, the decision was made for the survey to slide in from the right of the screen, covering half of the screen with a message to ask the nurse to fill in the survey. This meant that it was necessary to also design a starting screen for the survey, rather than immediately showing the survey. This starting screen would allow clicking away the survey, hence allowing the nurse to opt out in case they did not want to fill in the survey, or if the nurse was too busy

at that moment. This would also allow the researcher to accurately be able to calculate the survey response rate.

For the starting screen, it would be important to already implement the BE principles, as this is a crucial point for a nurse to decide whether to take the survey. So, four versions of the starting screens would have to be made for each of the three principles and the version without BE. In addition, the starting screens should have a similar layout in order to avoid other factors than BE influencing the survey response rate. As seen in Appendix J – Survey Starting Screens, the starting screen has an image at the top. This image was necessary to bring some BE principles more alive. With authority bias, it became important to clearly show the organization as the authority figure. So, the logo of Health Corp was included here. For social influence, the nurses had to be influenced by the other nurses around them. To highlight the part of other nurses, a picture of a nurse was used. However, for the loss aversion version of the survey, no picture was found that could accurately embody this principle without confusing the nurse. So, the decision was made to use the logo of the application for both loss aversion and the non-BE version.

Then, in the text, the nurse would be asked to take part in the survey. Here, too, the BE principles had to be included. For authority bias, Health Corp was mentioned to be the one asking for the feedback given in the survey and that Health Corp would be appreciating it. For social influence, it was highlighted that other nurses had already been answering the survey, nudging the nurse in the same direction. Loss aversion made it clear that the nurse would be losing the opportunity to share their opinion on changes to make the application better. Finally, the non-BE version would simply ask the nurse to fill in the survey.

The final element of the starting screen was that of the button agreeing to take part in the survey or to opt out of answering the survey. The conscious decision was made to make the accept button more prominent, which was something that had been noticed to be done more often in other surveys. The 'positive' answer would be clearer than that of declining. This was done in multiple ways, such as greying out the decline button while the accept button would be a bright colour. However, the decision was made to make the opt-out option underlined text as this was a common practice at Health Corp. The final decision that was made here was to have a small difference in the phrasing of the decline option in the loss aversion version, which was to emphasize 'losing' something. So, instead of using the *'Not this time'* option, the decision was made to phrase it as *'No, I'll miss out'*. The result of such a starting screen design can be seen in Figure 18, with the other starting screens being shown in Appendix J – Survey Starting Screens.
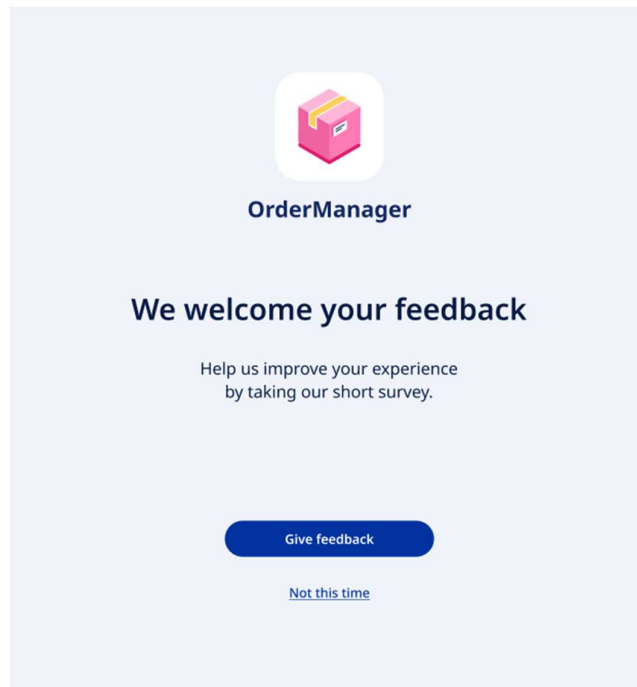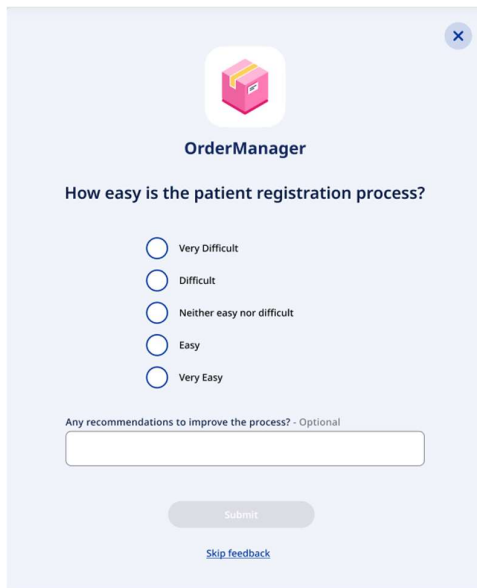
Figure 18: One of the four designs created for the starting screen of the survey.

If the nurse would accept the survey on the starting screen, then the questions regarding the process would be asked. As mentioned, these questions were previously decided upon and would be implemented in this field. The decision was made to only carry over the same images relating to the different versions from the starting screen into this field as it was deemed to be rather difficult to implement the BE principles into the questions while keeping the questions short and straightforward. The reason to keep the questions short and straightforward is that literature had shown that it is best to create these types of questions to receive a response (Qualtrics, 2022). An example of what this would look like is shown in Figure 19 and the other designs can be found in Appendix K – Survey Questions.

Finally, to formally close the survey it was decided to create an ending screen to thank the nurse for filling out the survey and submitting this. In order to keep the thank you screen from looking blank, an image was added which showed that the feedback given was sent off to Health Corp. This decision was made in consensus with the designers to make it more prominent to the nurse that their feedback would be received and used for further decision-making within Health Corp. This same message would be supported by mentioning how the feedback would be used to improve the application. Also, to make it more personal the name of the nurse was added to the thank you message. The final design of this screen can be found in Figure 20.

Across all the designs, the colour scheme that was used was that of Health Corp itself, which was blue and grey.

Figure 19: One of the four designs created for the questions of the survey.
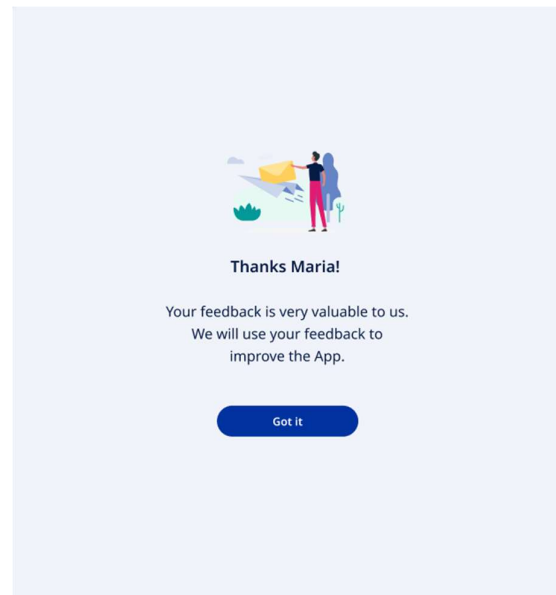

Figure 20: The ending screen design for the surveys.

In-App Survey Logic

After finalizing the designs of the surveys and each of their versions, a next decision had to be made on when these surveys would show up and how often they would be showing up. The designers had already decided that a survey for a specific process should only show up once per quarter for an individual nurse. However, this logic had to be refined further to fit the research. So, instead, a survey would only pop up once for one process, only once per individual, and only once per day until the nurse would have submitted a survey. So, this meant that if the nurse would answer the survey for a process, then the nurse would not be asked again to fill in a survey. On the contrary, if the nurse would decline to participate in the survey for a process, then the nurse could get asked to answer a survey for the same process again the next day.

To be able to test each version of the survey fairly, this meant that the versions had to be equally distributed. A meeting was held with multiple developers to discuss how this could be done. Things that were taken into consideration were workload and complexity. After thinking of multiple ways to do this, for example by splitting the entire user base into four, one section representing one version of the survey, the final decision was made to alternate between the version upon each popup. According to the developers, this was easier to implement on their side while it would also keep the distribution more equal. The reasoning behind this is that if nurses representing a version of the survey are more active than other groups, the outcome of the survey response rate could be influenced. The alternation of surveys meant that the first version that would show up would be version A, the next version would be version B, and so on. Upon reaching the fourth version (version D), the next version would start at version A again.

Data Collection

After finalizing the question formatting, the survey designs and the logic behind the survey showing up, a decision had to be made on which data should be extracted from the survey. The data that Health Corp deemed important first had to be discussed with the supervisor at Health Corp. The researcher held a meeting with the supervisor to find the most important data for Health Corp. Following this meeting, a lot of data topics were mentioned, such as process name, rating, date created, user e-mail, user mobile, user landline, clinic name, and clinic SAP ID. However, with this personal information to be tracked, the legal department had to get involved which will be further discussed in the next paragraph. Following multiple meetings with the legal department and careful consideration of the necessary information for Health Corp and this research, it was decided to track the following: the name of the process that the survey was triggered in (the point of conversion), the rating that was given in response to the question, the platform that was used to respond to the survey (either from the mobile application of a patient, or through the mobile or web version of the application of the nurse), the chosen option from a set of multiple choice options to answer the question with, comments regarding the given answer, the country in which the survey was answered, the type of user (either a nurse or a patient), the mode that the user was in (the trial mode of the application, or the production mode with full functionality), the date that the survey was submitted, the version of the survey that was answered (either the treatment or the control group), and finally the survey response rate.

In addition to this, a place to store all this data had to be decided upon. To create a report extracted from the SQL database where all the raw data would be stored, it was suggested to create a report in Excel with the necessary data. This way, the information stored in Excel could easily be shared with the developers, designers, the supervisor and the researcher. To share this information, the Excel document would be sent out with the updated data every Friday morning. This document can be found in Appendix M – Excel Report Survey Data.

Privacy Policy Adjustment

As mentioned in the previous paragraph, some of the data that was deemed personal became an issue with the privacy policy at the time. To discuss this topic, a meeting was held together with a senior director from the legal department and one of the designers. In this meeting, the senior director was given information about the goal of the survey and which data was intended to be gathered from the survey. During this meeting, the senior director mentioned that the privacy policy did not cover the data that was mentioned here. In addition, the senior director suggested keeping any collection of personal data to a minimum and only collecting strictly necessary data to ensure the end user's privacy and experience. So, the data that was deemed unnecessary to collect was removed from the list and the finalized data collection list that was mentioned in the previous paragraph was. Following this meeting, the senior director sent out an email with specific questions to be answered in the next meeting so that the privacy policy could be adjusted in such a way that it would cover all the personal data that would be collected. The questions that were asked were as follows:

- Which legal entity will act as the data controller and decides to implement the survey?
- Is it possible to achieve the objective without the collection of personal data?

- What personal data will be collected?
- What will the data be used for?
- Where will the data be hosted?
- Who will have access to the data?
- How long will the data be stored?

To cover these questions, a follow-up meeting was held where all these topics were covered by showing the process of how the data would be collected, reported and evaluated. This process can be found in Figure 21. Once it was clear to the senior director what had to be changed in the privacy policy to allow for the survey to run, the senior director made these changes and contacted the researcher, supervisor, and designer to inform them about the changes and that the survey was allowed to be implemented.

Figure 21: Data collection process as explained to the legal department at Health Corp.

4.3. Action Taking

Following the finalization of the survey design and the adjustment of the privacy policy in the action-planning phase, the action-taking phase could begin. In this phase, the survey was implemented into the web application of Health Corp by the developers. Upon implementation and its release, the nurses were able to take part in the surveys and the data collection could start and be interpreted.

In-App Implementation

This research is using an A/B testing approach to test the performance of multiple different BE principles. However, to get an accurate measurement of the performance each version created to represent a BE principle had to be implemented at the same time to attempt to avoid external factors interfering with the performance. So, each survey design had to be implemented at the same time so that these could run simultaneously under the same circumstances.

To start this implementation, the survey designs that had been created by the designers, which can be found in Appendix J – Survey Starting Screens, Appendix K – Survey Questions,

and Appendix L – Survey Thank You Screen, had to be handed over to the developers to be able to put these into the application as intended. Then, the logic that had been defined had to be communicated in detail to the developers so that this logic could be implemented into the system that allows the application to run. Both tasks were created in the sprint planning of the week of the 26th of September and finished by the 7th of October. So, on October 10th the survey was active in the application. Upon this, the data could be passively collected upon submission of the surveys.

Data Collection
Once the survey was implemented in October, the survey received only 8 responses by the end of the month. However, as the survey continued to run in November, another 25 responses were received. To receive as many responses as possible, the survey continued to run in December and ended up receiving another 19 responses, summing up to 52 responses by the end of the year 2022. In 2023, an additional 42 responses were received which have been included in the results of this research.

Results
After running the survey from October through January, a total of 94 responses were received. Among these 94 responses, 27 responses were to the social influence survey version, 19 to the loss aversion survey version, 10 to the authority bias survey version, and 38 to the non-BE survey version. Please see Table 2 for the overview of these results. Surprisingly, the survey version without BE was able to receive the highest response rate of 25%, with authority bias and loss aversion receiving a response rate of 9% and 11% respectively. Social influence received the lowest survey response rate at 7%.

Table 2: Experiment results for the survey response rate.

Survey Version	Survey Requests	Survey Submissions	Response Rate
No BE	150	38	25%
Authority Bias	111	10	9%
Social Influence	381	27	7%
Loss Aversion	179	19	11%

4.4. Evaluation
In this section, the result of the experiment will be discussed. Potential explanations for the differences in response rate will be supported through literature and case studies that used the BE principles that were used in this research but for another means.

4.4.1. Results
After concluding the experiment, the final results of the performance of each version were found. In Table 2, it can be seen that the lowest survey response rate is that of social influence, followed by loss aversion and authority bias respectively, with finally the version without BE having the highest survey response rate. This result indicates that the BE principles that have been tested in this experiment (authority bias, social influence, and loss aversion) do not have a positive impact on the response rate of surveys. This is a surprising result, as the literature suggests that involving BE in design is potentially applicable in situations where human

behaviour is specific, such as increasing survey response rate (Meija, 2021). So, the question is why the result contrasts the expectation of the experiment. To do this, Pearson's Chi-Squared Test will be conducted to compare the observed results against the expected results.

4.4.1.1. Pearson's Chi-Squared Test

From the data that was gathered from the survey, a chi-squared test is to be done to further investigate the data. Below in Table 3, the data that was gathered is shown. For each version of the survey, the number of nonresponses and responses to the survey are shown, as well as the total number of nonresponses and responses. Version A of the survey corresponds to the no-BE version of the survey (the control group). Versions B, C, and D correspond to the authority bias, social influence, and loss aversion versions of the surveys (the treatment groups). Finally, the total sample for each survey version as well as the overall sample is shown. The bottom row shows the survey response rate, which is where it can be seen that version A, which is the control group, received the highest survey response rate.

	A	B	C	D	Row Total
Nonresponse	112	101	354	160	727
Response	38	10	27	19	94
Column Total	150	111	381	179	821
Response Rate (in %)	0,25	0,09	0,07	0,11	

Table 3: Observed values for the chi-squared test.

To further investigate the data, some hypotheses must be made to make a conclusion about the expected values. There are two hypotheses for the expected values:

- Null hypothesis: the survey version does not have an effect on the survey response rate.
- Alternative hypothesis: there is a relationship between the survey version and the survey response rate.

Based on the observed values in Table 3, the degree of freedom (df) can be calculated. In this formula, r corresponds to the number of rows, while c corresponds to the number of columns. After the calculation, the degree of freedom is equal to 3.

$$df = (r - 1)(c - 1)$$
$$df = (2 - 1)(4 - 1)$$
$$df = (1)(3)$$
$$df = 3$$

To calculate the expected values, the row total will be multiplied by the column total and divided by the sample size. This calculation is done for each survey version and every nonresponse and response row. The final result can be seen in Table 4.

	A	B	C	D	Row Total
Nonresponse	133	98	337	159	727
Response	17	13	44	20	94
Column Total	150	111	381	179	821
Response Rate (in %)	0.11	0.11	0.11	0.11	

Table 4: Expected values for the chi-squared test.

In Table 4, all the response rates are equal to 11%, which would be expected if the null hypothesis is supported. The observed values from Table 3 can be compared to the expected values from Table 4 to see the difference between them. If the difference is zero, it indicates that the null hypothesis is valid. However, when comparing these values, version A has a lower expected response amount compared to the observed values. This suggests that people responded more often to survey version A than expected. On the contrary, versions B and C had a lower response amount than what was expected. Meanwhile, version D had a nearly similar response number in the observed and expected values. To mathematically compare the observed and expected values, the chi-squared test will be used.

Pearson's Chi-Squared Test uses the following formula:

$$x^2 = \sum \frac{(O - E)^2}{E}$$

The squared difference is calculated for each combination of survey version and response type, so nonresponse or response. The result of this can be found in Table 5.

Survey Version	Response	Observed	Expected	Squared difference / Expected
A	Nonresponse	112	133	3.32
A	Response	38	17	25.94
B	Nonresponse	101	98	0.09
B	Response	10	13	0.69
C	Nonresponse	354	337	0.86
C	Response	27	44	6.57
D	Nonresponse	160	159	0.01
D	Response	19	20	0.05
			SUM	37.52

Table 5: Pearson's Chi-Squared Test.

In Table 5, the chi-squared value is calculated to be 37.52. To further explain this number, the chi-square distribution table can be used which can be found in Appendix O – Chi-Squared Table. When looking at the chi-square distribution table and using the default significance of 0.05 and the calculated degree of freedom of 3, the critical value is 7.815. As the chi-square value is much higher than the critical value, the null hypothesis has to be rejected. This

indicates that the survey response rate is dependent on the survey version. So, based on the observed data, the conclusion can be made that participants respond better to survey version A, than any of the other surveys.

4.4.1.2. Response Rate Analysis on Country and Process Level

Aside from the survey response rate only specifically of each survey version, it may be interesting to see if the specific countries or the specific processes had anything to do with this. Using the ratio of the overall requested surveys and the requested surveys per country and requested surveys per process a response rate has been defined for each country (Germany, Sweden, and Spain), and each process (order creation, patient creation, and prescription creation). Germany received the most requests because the survey was first rolled out here and has the most amount of nurses using the application. The ratio for Germany is 9:10. Spain was added later and has a ratio of 2:25. Lastly, Sweden was added towards the end of the research and received only 14 survey requests in total, meaning that it only received 1% of the total requested surveys. So, the results of Sweden are not likely to be reliable enough to attach an analysis to. The survey response rates in Table 6 are the result of this analysis.

Table 6: Response Rate analysis per country and process.

Survey Version		Germany	Sweden	Spain	OrderCreation	PatientCreation	PrescriptionCreation
No BE	**Requests**	135	3	12	83	35	32
	Submissions	37	0	1	22	2	14
	Response Rate	27%	0%	8%	26%	6%	44%
Authority Bias	**Requests**	100	4	7	62	26	23
	Submissions	5	4	1	8	1	1
	Response Rate	5%	100%	14%	13%	4%	4%
Social Influence	**Requests**	347	4	30	213	88	80
	Submissions	25	0	2	20	3	4
	Response Rate	7%	0%	7%	9%	3%	5%
Loss Aversion	**Requests**	162	3	14	100	41	38
	Submissions	13	0	6	7	5	7
	Response Rate	8%	0%	43%	7%	12%	18%

In Table 6 it can be seen that Germany gave the highest response rate for the survey version of the control group (no BE) at 27%, which is 2% higher than the overall survey response rate given to the control group which is what is meant by overall survey response rate unless specifically named otherwise. When comparing the survey response rates given to the other versions of the survey, that of the control group is much higher. Interestingly, while the response rates to the loss aversion and social influence versions of the survey are similar to

that of the overall response rate, the response rate of Germany to the authority bias is much lower than that of the overall response rate.

Sweden received too few requests to deeply explore but can be briefly discussed. No responses were received for the control group version of the survey, nor social influence or loss aversion. This is why Sweden has a 0% survey response rate for these 3 versions. However, taking the ratio of Sweden across the different survey versions, it can be said that Sweden has given a 100% response rate to the authority bias survey. Still, as the responses are not enough nor are the requests, the data is not sufficient enough to provide an explanation.

The final country among the three is Spain. Spain gave the highest response rate for the loss aversion version of the survey, which is significantly higher than that of Germany and also that of the overall response rate. The social influence and authority bias versions of the survey are very similar to that of the overall response rate, so there is no surprising finding here. However, it is important to note that the control group version of the survey performed far worse in Spain than it did in Germany as well as in the overall response rate.

Now, the response rate of each process across all the survey versions will be explored. The first process is the order creation process. The survey response rates across all versions of this process are close to that of the overall survey response rate. The only positive difference is that for the social influence version, with a difference of 2%. However, as there are no significant differences between the process and the overall, no interesting findings can be made here.

The second process is the patient creation process. Immediately it can be seen that there is a large difference between the response rate for this process and the overall response rate when it comes to the control group version of the survey. Where the overall response rate is 25%, the response rate for this process on this version of the survey is only 6%. A larger difference in survey response rate between these two can also be found in that of authority bias. Social influence and loss aversion have a similar survey response rate to that of the overall.

The last process is the prescription creation process where there is a large difference in the survey response rate of the process itself and the overall survey response rate. Where the overall survey response rate is 25%, that of the prescription creation process is at 42% which means that nearly half of the sent-out surveys for the control group were responded to in this process. While the survey response rate to social influence was nearly identical to the overall survey response rate, the survey response rate to the authority bias was much lower. Also, the survey response rate in this process to loss aversion was almost double that of the overall survey response rate. So, all survey versions in this process except for social influence differed in their response rate compared to that of the overall response rate.

In conclusion, when analysing the survey response rate specifically for the control group version in comparison to the survey overall response rate, the biggest difference is that in the country Spain and the processes of patient creation and prescription creation. Where the survey response rates of the control group version for Spain and the patient creation process were much lower in comparison to the overall survey response rate, the survey response rate

was much higher for the prescription creation process. For the authority bias version, the survey response rates for Germany, excluding Sweden due to insufficient data, and for the patient creation and prescription creation processes were much lower than the overall survey response rate. Then, in the social influence version, the survey response rate was similar to the overall survey response rate across all countries and all processes, only having a slightly higher survey response rate for the order creation process which is not likely to result in an interesting finding. Finally, the loss aversion survey version was found to have a large difference in response rate, especially for Spain, but also for the prescription process. Spain quadrupled the survey response rate to this version and the survey response rate in the prescription process was doubled in comparison to the overall survey response rate.

4.4.2. Validity Considerations

A study conducted by Holtom, Baruch, Aguinis, and Ballinger (2022) provides a framework to evaluate the validity of the survey response rate. This framework consists of multiple questions to be answered, for example asking if the participant is receiving a payment for completing the survey. For ease, the questions and the answers are found in Appendix D – Validity Assessment Framework. With this framework, the authors attempt to address that a low survey response rate does not necessarily mean that the quality of the data is low and vice versa. The result of this evaluation process is that regarding the validity of the result, the respondent might have felt coerced to answer the survey as the respondent is in a business-related relationship with Health Corp, thus lowering the validity. However, the survey is significantly valid concerning the participant qualifications as the respondent is sufficiently qualified for the survey. In addition to this, as participants were not out to respond just to receive a payment the quality and validity of the research are evaluated to be appropriate.

Additionally, as this research has only been conducted with a specific group of participants, it is important to stress that the results that are received are group-specific results and may not apply to other groups falling outside of this segment (Bojinov, Saint-Jacques, & Tingley, 2020). Therefore, the results should have a high external validity for the specified target group, but as it is not generalizable across a larger population it may not have a high external validity for this purpose. A common mistake that is faced in A/B testing is that the tests are ended too quickly once a 'clear' winner is found among the groups (Gallo, 2017). This type of mistake tends to impact the internal validity of the experiment. However, this research has run for a bit over two months, ensuring enough time to naturally move towards a high internal validity. So, although the external and internal validity of the research is not likely to be low, the external validity for further generalizability is low.

4.4.3. Meta-Analysis on Behavioural Economics and Survey Response Rate

This section shows the result of a meta-analysis on the relationship between behavioural economics and survey response rate.

4.4.3.1. Introduction

Rationale & Objectives

To increase the low survey response rate of online surveys, BE was found to be a potential solution to boosting the survey response rate. However, the result of the experiment in this thesis resulted otherwise and the collected data was insufficient for data analytics. A

recommendation on the usage of BE to improve the survey response rate should be based on more evidence to prove the legitimacy of the result of the experiment. Therefore, a systematic review was conducted. The objective of the review was to evaluate the results of other published studies about the effectiveness of BE on survey response rate and to finally compare these results to the result of this thesis.

Design

This study uses a systematic review and meta-analysis of multiple studies where the use of BE and its effect on survey response rate was researched. To ensure appropriate reporting of the systematic review and meta-analysis, the Preferred Reporting Items for Systematic Review and Meta-Analysis 2020 (Page, et al., 2021) will be followed.

Procedure

Following the PRISMA guidelines, the overview of topics for reporting was made. This overview includes defining the eligibility criteria (of articles), the information sources, the search strategy, the selection process, the data collection process, a study risk of bias assessment, measures, synthesis method, reporting bias assessment, and certainty assessment.

4.4.3.2. Method

Eligibility Criteria

Articles were included if: (1) the population studied were exposed to online surveys; (2) BE principles were used as interventions to improve survey response rate, even if the BE principles are not used in this thesis; (3) the result of the effectiveness BE on survey response rate was reported; (4) original data was reported in the article, and (5) the article was written in the English language.

No restriction was set on the publication year.

Information sources

To find the articles, Google Scholar was searched. The database had been searched between February 2022 through May 2023. The last search was done on May 18, 2023.

Search Strategy & Selection and Data Collection Process

The literature search was conducted in Google Scholar. The keywords that were used in the search were: ("behavioural economics" OR "behavioural science") AND ("online survey" OR "web survey" OR "internet survey") AND ("response rate" OR "completion rate" OR "participation rate") AND (impact OR effect OR influence OR efficacy OR effectiveness). The initial decision on the eligibility of the article was made based on the title of the article. If it passed, then the abstract was scanned to ensure relevancy. Finally, each eligibility criterion was consulted by the researcher to determine the final decision on the eligibility of the article.

As the research, as well as the review, was conducted by one researcher, this researcher reviewed all articles found in Google Scholar. Therefore, the single screening is efficient but is sensitive to missing relevant articles (Page, et al., 2021). No automation tools, crowdsourcing or known assessments were used to collect data.

Risk of Bias Assessment

An easy method and tool developed by McGuiness & Higgins (2021), called ROB 2.0, was used to assess the selected articles on their bias. Based on the evaluation of their bias, the decision was made to retain the articles or reject them.

Measures

The main outcome of the study was if a BE intervention in an online-conducted survey had an effect on the survey response rate. Exploratory outcomes were not considered for this study. All types of BE interventions were included in the search. The eligible outcomes were categorized as:

- Survey Response Rate
  - Response rate
  - Completion rate
  - Participation rate

Any measure for survey response rate was eligible. All outcomes were measured through online surveys.

Synthesis Method & Reporting Bias Assessment

As many different BE principles can be implemented in different intervention contexts in an attempt to improve the survey response rate, a process was created for deciding the eligibility of a study. To decide its eligibility, the following categories were made: (1) if a BE principle was used as part of the intervention; (2) if the intervention was included in an online conducted survey setting; and (3) if the comparison of survey response rate was included in the results of the study.

The analysis was done using the RevMan Web tool, using its modules for meta-analysis. In this analysis, the effect of BE principles as interventions on improving the survey response rate was considered. To do so, the data of the studies was extracted, and the risk ratio using the default confidence interval of 95% was calculated and visualized in a forest plot. Therefore, if the study received a risk ratio of below 1.0 it indicates that the BE interventions increased the survey response rate, while receiving a risk ratio of 1.0 would indicate that the BE intervention has no effect. Finally, if the study received a risk ratio of over 1.0, it would indicate that the BE interventions would decrease the survey response rate.

To assess risk due to publicity bias or missing data, a funnel plot was used to detect asymmetry.

4.4.3.3.   Results

Study Selection & Characteristics

A total of 7020 articles were found using the advanced search option in Google Scholar with the relevant keywords. As the researcher is doing the systematic review by herself without the use of tools, the articles had been sorted by relevance and the first 100 articles were chosen to be screened. During this screening, another 85 articles were removed after reading their title and abstract as they did not match the topic of the research. The remaining 15 articles were sought for retrieval and 1 article was not able to be retrieved. The 14 articles

were then assessed for eligibility of which 8 articles were not deemed eligible (Daikeler, Silber, & Bosnjak, 2021; Jin, 2011; Nichols, Konya, Horwitz, & Raim, 2019; Edwards, et al., 2009; Connors, Krupnikov, Yanna, & Ryan, 2019; Lindgren, Markstedt, Martinsson, & Andreasson, 2018; Wen & Fang, 2012; Seshadri & Broekmier, 2022). These articles were omitted, because each of these articles did not report the response rate specifically, but either other data similar to the response rate or no response rate at all. Another reason was because the article was a meta-analysis so it did not show original data. A third reason was that the study used multiple groups, which would be difficult to work with later on in the review and to create a forest plot.
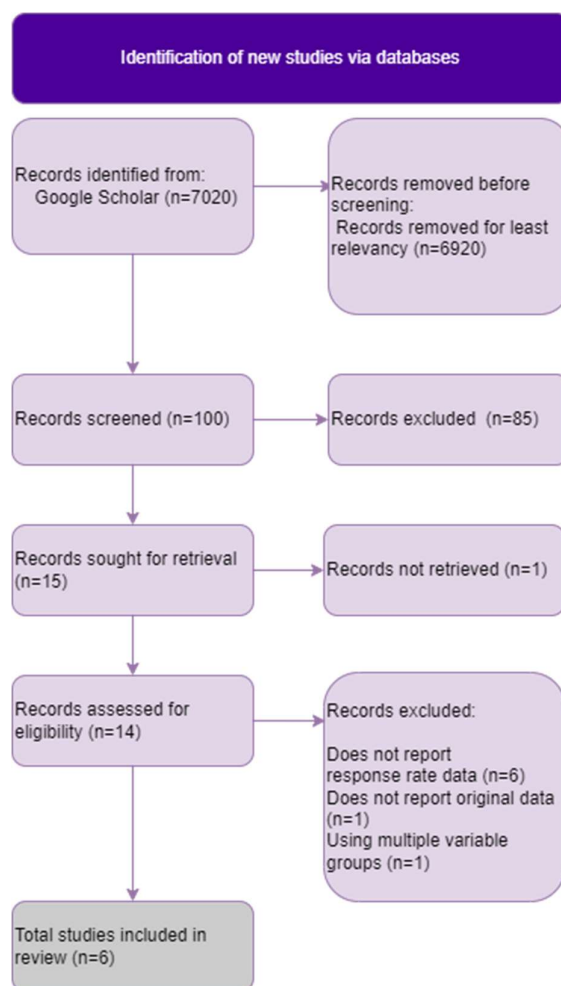
Figure 22: Study selection flow diagram.

So, 6 articles were left to be included in the review. This flow can be seen in Figure 22.

For these 6 studies, the effectiveness of BE principles on survey response rate was examined. Table 7 shows the studies that were included in the review and for each study the study design, country, sample size, the used intervention (BE principle) and the outcome (survey response rate) are shown.

Study characteristics				Intervention	Outcome
Author (year)	Study design	Country	Sample size	BE-principle	Survey response rate
Stanley, et al. (2020)	Quasi-randomized trial	USA	13162	Incentives	Increased
Hathaway, et al. (2021)	Quasi-randomized trial	USA	2750	Incentives	Increased
Cook, et al. (2016)	Randomized Controlled Trial	USA	3966	Incentives	No effect
Rackoff, et al. (2023)	Randomized Controlled Trial	USA	45469	Framing, Reciprocity, Inclusivity	Decreased
Mueller, et al. (2014)	Randomized Controlled Trial	Germany	815	Anonymity	No effect
Warwick, et al. (2019)	Randomized Controlled Trial	USA	3000	(Social) Incentives	No effect

Table 7: Selected studies' characteristics.

Risk of Bias

For the risk-of-bias assessment, ROB 2.0 was used. For the summarization, please look at Figure 23. When looking at the overall risk result, 4 of the 6 studies received some concerns about the risk of bias in at least one domain. However, the concerns were not too large for it to be considered a high risk of bias. Therefore, none of the studies had to be rejected, but the studies with some concerns should be carefully interpreted.

Domains:
D1: Bias arising from the randomization process.
D2: Bias due to deviations from intended intervention.
D3: Bias due to missing outcome data.
D4: Bias in measurement of the outcome.
D5: Bias in selection of the reported result.

Judgement
- Some concerns
+ Low
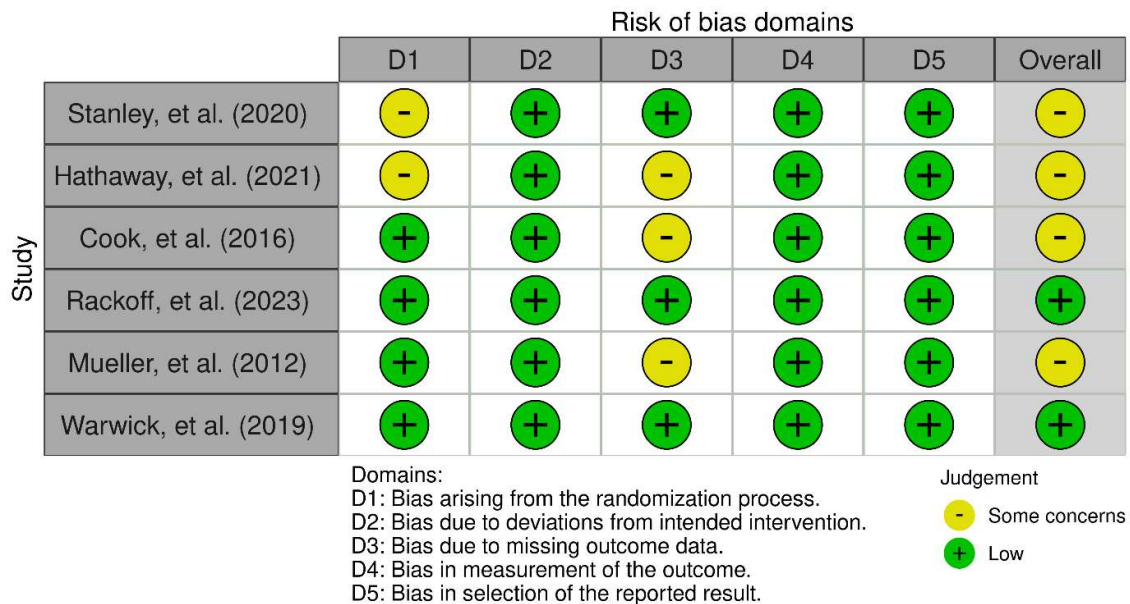
Figure 23: Selected studies' risk of bias evaluation.

Results of Individual Studies and Synthesis

As all of the studies are (quasi) randomized controlled trials, the RevMan Web tool is used to show the summary statistics for each of the studies. In Figure 24 the number of events and the total sample are shown for the group that was not exposed to a BE intervention and the group that was exposed to the BE intervention. The risk ratio and confidence interval for the dichotomous outcome of these studies are at the default of 95%.
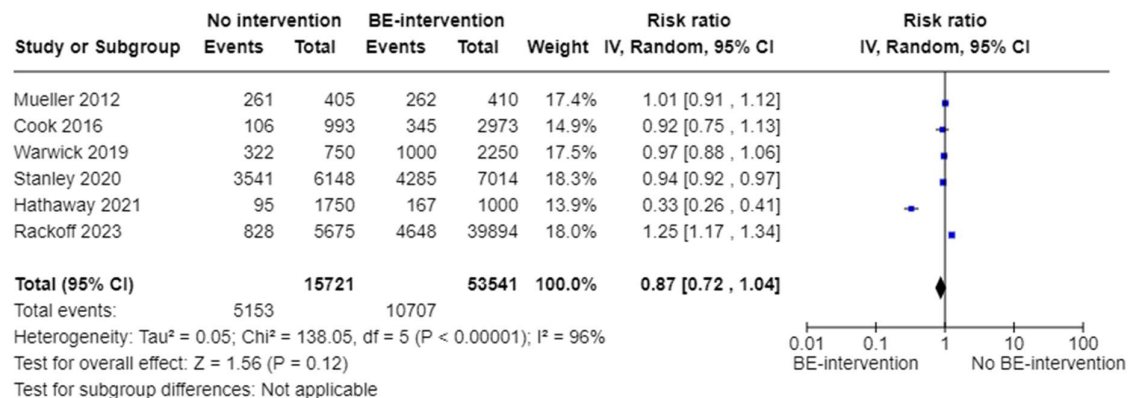
Study or Subgroup	No intervention Events	Total	BE-intervention Events	Total	Weight	Risk ratio IV, Random, 95% CI
Mueller 2012	261	405	262	410	17.4%	1.01 [0.91 , 1.12]
Cook 2016	106	993	345	2973	14.9%	0.92 [0.75 , 1.13]
Warwick 2019	322	750	1000	2250	17.5%	0.97 [0.88 , 1.06]
Stanley 2020	3541	6148	4285	7014	18.3%	0.94 [0.92 , 0.97]
Hathaway 2021	95	1750	167	1000	13.9%	0.33 [0.26 , 0.41]
Rackoff 2023	828	5675	4648	39894	18.0%	1.25 [1.17 , 1.34]
Total (95% CI)		**15721**		**53541**	**100.0%**	**0.87 [0.72 , 1.04]**
Total events:	5153		10707			

Heterogeneity: Tau² = 0.05; Chi² = 138.05, df = 5 (P < 0.00001); I² = 96%
Test for overall effect: Z = 1.56 (P = 0.12)
Test for subgroup differences: Not applicable

Figure 24: Summary statistics of the selected studies.

These six studies all analysed the effect of a BE principle on the survey response rate of an online survey. Across these studies, 5153 of the 15721 recruited participants were not exposed to a BE principle and therefore functioned as the control group, while 10707 of the 53531 recruited participants were exposed to a BE principle. Of these six studies, four studies had some concerns with the risk of bias, while the remaining two studies were a low risk of bias. Across the studies, different BE principles were used as interventions: incentives, framing, anonymity, reciprocity, and inclusivity. Four studies used incentives as the chosen BE intervention, but two appeared to have an effect on the survey response rate while the other two did not report an effect. However, it should be mentioned that the two studies reporting that incentives have an effect on survey response rate had some concerns with the risk of bias. For the studies that reported no effect of incentives on survey response rate, one study had some concerns with the risk of bias while the other study had a low risk of bias. The study that analysed the effect of framing, inclusivity and reciprocity on survey response rate reported that these BE interventions decreased the survey response rate. Additionally, this study had a low risk of bias. Finally, the study that reported on the effect of anonymity on survey response rate reported no effect on its outcome. However, this study had some concerns with the risk of bias.

When interpreting the studies in the forest plot individually, it shows that the studies by Mueller (2014), Cook (2016), and Warwick (2019) are not statistically significant as the confidence intervals fall on both sides of the null hypothesis. However, the studies by Stanley (2020), Hathaway (2021), and Rackoff (2023) are statistically significant as the former two studies are on the left side of the null hypothesis and the latter study is on the right side of the null hypothesis.

Across the 6 studies, which include a total of 69262 participants, the effect of BE principles on survey response rate has been reported. When looking at the combined statistics of all the studies, the summary effect is measured to be 0.87. However, when looking at the confidence intervals the range crosses the null hypothesis line. This indicates that the overall effect is not statistically significant. This is supported by the P-value given by the test for overall effect, which is 0.12 and therefore not statistically significant. Furthermore, when looking at the homogeneity of the study, I^2 is incredibly high at 96%. This indicates that there is a high heterogeneity between the studies in this review and that there is only a 4% variation which is because of chance.

Sensitivity Analysis & Reporting Bias Assessment

To find the cause behind the heterogeneity, a sensitivity analysis in RevMan Web has been done which is shown in Figure 25. After removing some of the studies from the analysis, it is found that the study of Hathaway (2021) and Rackoff (2023) reduce I^2 down to 0%. This indicates that these two particular studies differ greatly from the other studies in the review. Therefore, the main analysis is inconsistent as it is greatly affected by this variation. It is likely that the differences in BE interventions, such as using incentives or motivational messages in the forms of reciprocity, have caused this high heterogeneity.

Study or Subgroup	No intervention Events	Total	BE-intervention Events	Total	Weight	Risk ratio IV, Random, 95% CI	Risk of Bias A B C D E F
✓ Mueller 2012	261	405	262	410	6.5%	1.01 [0.91 , 1.12]	
✓ Cook 2016	106	993	345	2973	1.6%	0.92 [0.75 , 1.13]	
✓ Warwick 2019	322	750	1000	2250	7.6%	0.97 [0.88 , 1.06]	
✓ Stanley 2020	3541	6148	4285	7014	84.3%	0.94 [0.92 , 0.97]	
✗ Hathaway 2021	95	1750	167	1000	0.0%	0.33 [0.26 , 0.41]	
✗ Rackoff 2023	828	5675	4648	39894	0.0%	1.25 [1.17 , 1.34]	
Total (95% CI)		8296		12647	100.0%	0.95 [0.92 , 0.97]	
Total events:	4230		5892				

Heterogeneity: Tau² = 0.00; Chi² = 1.78, df = 3 (P = 0.62); I² = 0%
Test for overall effect: Z = 3.99 (P < 0.0001)
Test for subgroup differences: Not applicable

Risk of bias legend
(A) Random sequence generation (selection bias)
(B) Allocation concealment (selection bias)
(C) Blinding of participants and personnel (performance bias)
(D) Incomplete outcome data (attrition bias)
(E) Selective reporting (reporting bias)
(F) Other bias

Figure 25: Sensitivity analysis results on selected studies.

To further check for any reporting biases, the funnel plot was used in RevMan Web. This can be found in Figure 26. The figure shows an asymmetrical funnel plot with the studies with the smallest number of events, Hathaway (2021) and Cook (2016), at the bottom of the plot. The studies of Hathaway (2021) and Rackoff (2023) fall outside of the 95% confidence intervals. This suggests that the effect sizes of these studies are different than expected compared to the other studies and are considered outliers. The remaining are close to the overall effect line which suggests that these studies have higher precision. The asymmetrical shape of the funnel plot indicates that there potentially is publication bias present. This is likely due to the analysis missing certain smaller studies that would have larger effect sizes. This is not a surprise as from the 7020 articles that were found, only 100 studies were scanned due to a

time limit and heavy load. Additionally, the different types of BE interventions have likely caused this asymmetry.
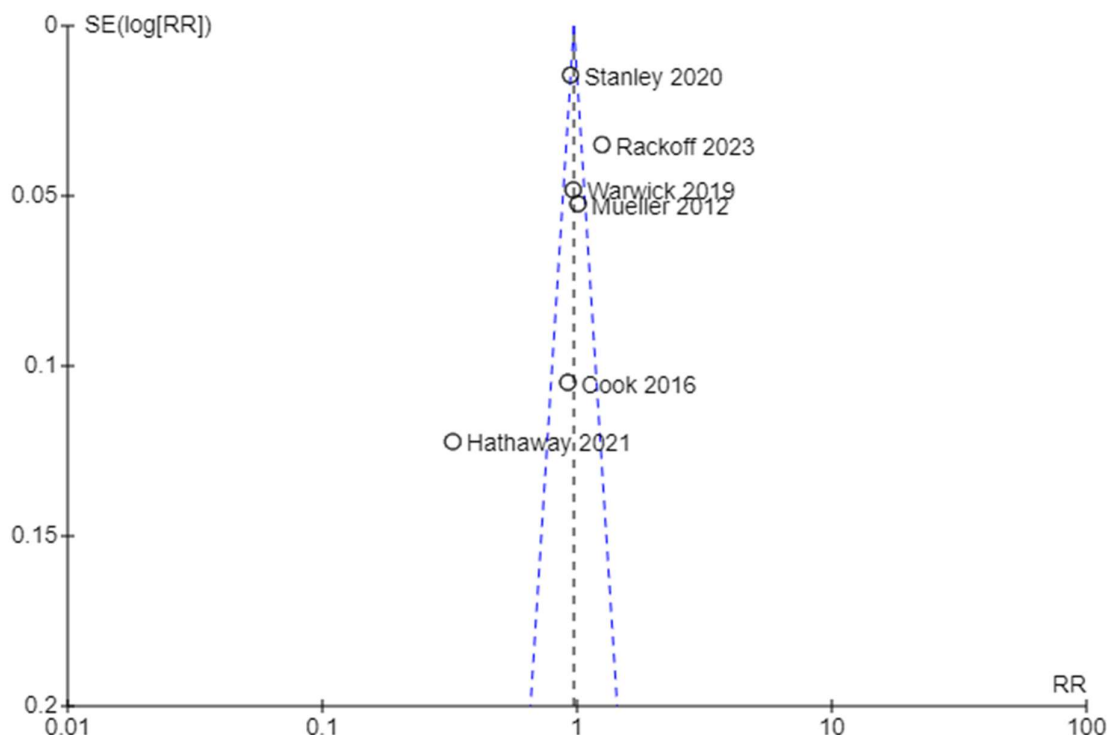

Figure 26: Funnel plot for detecting reporting biases.

4.4.3.4.   Discussion

This meta-analysis reviewed six studies, therefore it is important to be cautious of interpreting the result. The review indicates that there is much heterogeneity across the six studies, which can be seen in Figure 24. This indicates that there is a lot of variability between these studies which is not due to random chance. This could be because of differences in study design, outcome measurements, or interventions. As BE interventions do fall under one large umbrella, the interventions are not the same. For example, some studies used monetary incentives to get participants to respond to the survey while another study used motivational appeals to get participants to respond which can be seen in Table 7. Therefore, it is extremely likely that this high heterogeneity is due to the differences in interventions. This suggests that the effects on the outcome differ significantly based on the intervention that is used which has to be considered in the interpretation of the overall result. This leads to the necessity of conducting subgroup analysis for each type of BE intervention rather than looking at the effect of BE as a whole. Furthermore, this result indicates that future studies that conduct meta-analysis must focus on comparing the specific interventions, such as using motivational appeals, and their effect on survey response rate separately.

Additionally, in Figure 24 the overall effect of the studies has a Z-value of 1.56 and a probability of 0.12. As the Z-value is small and the probability of observing the effect by chance is high,

the overall effect is not statistically significant. Therefore, along with the small number of studies observed, it is also important to be cautious in interpreting these results as the effects may only be observed due to random chance. Although the overall effect aims to show that BE interventions have a positive effect on survey response rate, the evidence is not strong enough to support this. However, when looking at the funnel plot in Figure 26, we can see that the studies that fall outside of the confidence intervals are those of Hathaway (2021) and Rackoff (2023), indicating a low precision. When removing these two articles in the sensitivity analysis which can be seen in Figure 25, the Z-value and the probability change significantly. The result of the sensitivity analysis shows a Z-value of 3.99 and a probability of less than 0.0001. This suggests that the overall effect is statistically significant and that the null hypothesis can be rejected, which supports the effect of BE interventions on the survey response rate. The extremely low probability value indicates that it is unlikely that these results are observed by chance and that there is a true effect.

The limitations of the evidence in this study are that some of the studies that were part of the review had some concerns with bias. Therefore, it is important to exercise caution when it comes to the outcomes of these studies. Additionally, the high heterogeneity has also made it rather challenging to draw a definitive conclusion. This heterogeneity is likely due to the differences in BE interventions that were used across the studies. Also, some of the studies have focused on certain populations, such as elderly people, university students, or patients during the COVID-19 pandemic. This may cause some issues with the generalizability of the results. Finally, due to practical constraints such as time limit and the overall capability of the researcher, fewer studies were able to be included in the review which can have influenced the robustness of the results.

Finally, the implications for practice suggest that caution must be exercised in interpreting the overall result, because of the limited number of studies included in the review. It is not recommended for future research to continue on this topic. However, if future research is still done, the importance of subgroup analysis for different BE interventions should be highlighted which means that there needs to be a focus on specific interventions rather than BE interventions as a whole.

4.5. Reflection

In the reflection section, potential explanations for the differences in response rate between the versions will be supported through literature and case studies that used the BE principles that were used in this research, but for another means. The aim of this section is to conclude how useful BE is in designing a survey to boost the survey response rate.

4.5.1. Behavioural Economics and Influence on Response Rate

Authority bias, social influence and loss aversion were the BE principles used in this research as an attempt to increase the survey response rate. However, none of these BE principles were able to do this. BE is typically applied in fields such as health care, finance, public policy, and marketing amongst others. For example, in health care, personal reminders as a form of BE were used to see if this would boost the medication intake of the patient which it did (Hallek, Ockenfels, & Wiesen, 2022). In finance, the BE principle of goal setting had been used to help clients improve their financial planning positively (Eberhardt, Brüggen, Post, & Hoet, 2021). In

the field of public policy, the social norms principle was applied in an attempt to reduce energy usage (Alcott, 2011). Finally, the BE principle known as rewards was used to prove that using a reward system makes it more likely for a customer to buy a product more quicker than usual and to retain customers (Kivetz, Urminsky, & Zheng, 2006). As BE does work across different industries, perhaps it is the question of how BE was applied in this research that did not work. To apply BE to the surveys, the focus has been put on the UI design of the surveys. However, when looking at case studies and research where BE has been applied, a wider aspect has been taken into account. The following sections will discuss some aspects that have been covered in BE-related studies, but have not been taken into account in this research and could have had an influence on the effectiveness of the BE principles.

In-Person Exposure

A well-known study to observe the effect of authority bias was done in 1963 by Stanley Milgram. In this study, the participants were split into two groups where one group would have to punish a mistake made by the other group through increasingly worsening electric shocks (Milgram, 1963). In the case of refusal, the researcher, who is the authority figure in this scenario, would order the participant in person to continue with the experiment. Each shock would increase by 15 volts, starting from 15 up to 450 volts. As a result of being encouraged by the authority figure, all participants kept going until 300 volts, with 65% of the participants reaching the maximum of 450 volts. So, this was a method where the BE principle was exposed to the participant in person.

Another study where a BE principle was exposed to the participant in person, was where social influence was used by a hotel to increase towel reusage and decrease water usage. This was done by writing on notice cards that are placed inside hotel rooms how a percentage of hotel guests reuse their towels (Shang, Basil, & Wymer, 2010).

Finally, another study where BE was exposed to the participant in person was a study conducted by Novemsky & Kahneman (2005). Among the 8 experiments done in this study, in one experiment the participants were given a mug and a questionnaire which asked the participant whether or not they preferred the mug in front of them or the amount of money stated on the questionnaire which ranged from $0.50 to $25.00.

These are some of the studies conducted with the BE principles used in this study. As each of these studies was successful in proving the effectiveness of the BE principle in question, it could indicate that for the BE principles to be effective it must be done in person to increase its intensity. In addition to this, one of the pioneers of BE, Richard Thaler, suggests that in-person interactions leveraging BE are beneficial for changing behaviour (Thaler R. , 2016).

Duration of Exposure

In several studies, the effectiveness of BE was amplified through its persistent exposure. The Milgram experiment shows how the participants were constantly exposed to the authority bias principle (Milgram, 1963). Throughout the entire experiment, the authority figure was present and enforced authoritative pressure on the participant to obey.

In a study where the goal was to reduce energy consumption, the researcher sent out reports monthly. In these reports, the energy usage of the household was compared to that of their

neighbours and was rated as either 'great', 'good', or 'room for improvement' (Costa & Kahn, 2013).

To prove the effectiveness of loss aversion, a study was conducted by Bećirović, Suhonjić & Stanić (2022) to use loss aversion to improve the engagement of students in a classroom. The control group received points for completing assignments, starting from zero. The treatment group started from 100 and got points deducted for not completing assignments. This was done across 15 weeks with the students receiving an update on their points weekly per e-mail. The students showed a stronger reaction to having their points reduced, rather than added.

These case studies show that repeated exposure to the BE principle suggests more effectiveness. According to the MINDSPACE framework that was used in this research this explanation is reinforced as the authors argue that repeated exposure to the BE principle supports the shaping of behaviour (Dolan P. , et al., 2011).

Participants' Awareness and Cognitive Load

A more recent version of the Milgram experiment is a study conducted by Burger (2009) who had made some changes to the experiment to make it more ethical compared to the original experiment. The participants were recruited through advertisements and made time free to participate in the experiment, but avoided a thorough briefing to prevent bias.

In a social influence study by Nolan, Schultz, Cialdini, Goldstein & Griskevicius (2011), the goal was to gather the participants' reasons for conserving energy while being exposed to messages levering social influence. The researchers had chosen multiple neighbourhoods and gave the participants the opportunity to opt out of the research. Over half of the households voluntarily participated in the study.

In a study conducted by Fryer, Levitt, List & Sadoff (2012), loss aversion was studied in teachers who were randomly recruited at 9 different schools. The experiment was introduced to them and they were given the opportunity to opt out of the experiment before a specific date. Of all these teachers, 52% participated in the experiment. With the goal to improve students' performance, which is part of a teacher's job, no additional cognitive load is considered as the research is part of work tasks and workflow is not interrupted.

In this thesis, the participants of the study were perceiving high work pressure, which is common among healthcare workers, and were unaware of taking part in this research as it was anonymously done. Additionally, participation in the research was not part of any work tasks and may have interrupted workflow. It is proven that high work pressure impacts an individual's decision-making process because of time constraints (Payne, Samper, Bettman, & Luce, 2008). Additionally, studies have shown that these workflow interruptions have multiple negative effects, such as decreased performance, an increase in error in the healthcare industry, but also emotional exhaustion and increased levels of irritation (Baethge & Rigotti, 2013; Elfering, Grebner, & Ebener, 2015; Pachler, et al., 2018). Therefore, the nurses may have been less likely to participate in the survey and instead clicked away from the survey out of irritation. Consequently, the results of this research may have been impacted negatively by not carefully considering this aspect.

Behavioural Economics in Design

A case study where authority bias was used to alter user interface, was a case where an electronic health record provider used dark patterns to convince clinicians into accepting the recommendation for the use of opioids as a form of treatment for patients (Capurro & Velloso, 2021). This was done by showing that the options in a follow-up plan for the patient were originally taken from an academic publication. Although this is a dark pattern form of authority bias, it does show how the design of the interface is more intensely showing authority bias as compared to this research, where this was attempted more subconsciously by showing the logo of Health Corp. Therefore, the BE principle may not have been potent enough to have an effect on the behaviour of the nurses.

In a study to see how social influence affects customer responses to advertisements, Baksy, Eckles, Yan & Rosenn (2012) used consumers' real peers to create a link between the brand and their social network. However, in this study, only a static and not real number of participants were shown to the nurses as participants who had already responded to the survey. So, in actual conversations between nurses, this number for example can have been disproven if the nurses mentioned that they had not responded to the survey.

A case study of where loss aversion is used in the design is for example Netflix, where the front page of the website shows clearly that a visitor without a subscription is able to have a free subscription for a month (Gearson, 2019). Multiple businesses do this as an attempt to keep retain the visitor and have them purchase a subscription once it runs out. The difference between this case study and the survey design that was used in this research is the intensity and emphasis that is put on the loss factor. In the case study, something is given to the participant first and then taken away, while in this research something much less tangible is given to the participant and is easier to be declined.

Cultural Considerations

As the survey had been rolled out in Germany and Spain (excluding Sweden), there is a possibility of cultural differences influencing how authority is perceived. According to Hofstede's model of national culture, Germany has a low power distance while Spain has a medium power distance (Hofstede, 2001). This means that Germany considers authority figures more equal to themselves, while Spain is more accepting of the power inequality existing. When looking at the survey response rate to this version in Germany and Spain, this seems to be an appropriate explanation. Where the survey response rate is about half of the overall survey response rate, 5% to 10%, the survey response rate in Spain is nearly identical to the overall survey response rate. So the cultural difference is very likely to have an impact on how this affects the survey response rate at 11% with Germany responding worse to authority bias while Spain has a small positive response to it. Another study confirmed that Germany tends to respond worse to nudges (Reisch & Sunstein, 2023).

According to a study by Wang, Rieger & Hens (2017), three clusters of Hofstede's cultural model impact loss aversion. These three are power distance, individualism, and masculinity. This means that a high score on each of these three clusters indicates a higher loss aversion. The result of these three clusters for Spain and Germany can be found in Table 8.

Table 8: The three clusters of Hofstede's cultural model for both Germany and Spain. Source: (Clearly Cultural, 2022)

	Germany	Spain
Power Distance (PDI)	35	57
Individualism (INV)	67	51
Masculinity (MAS)	66	42

At first glance, this would indicate that Germany is more loss averse than Spain, even though Spain had a much higher response rate to the survey for the loss aversion survey than Germany (42% to 8%). Additionally, the result of cultural comparison in this same study by Wang, Rieger & Hens (2017) between Germanic cultures and Latin European show a similar loss aversion. However, when looking at the median of the countries, Spain does score a little higher than Germany (2.4 to 2.0). Furthermore, a study from 2007 conducted by Gächter, Johnson & Herrmann showed that a higher household income is related to a higher loss aversion (Gächter, Johnson, & Herrmann, 2007). When we compare a nurse's salary to the average salary in their country, we can see that Spanish nurses appear to have a higher income than Germany. Where in Germany the yearly average salary is €49,200, a nurse only makes €33,654 a year which is much lower (Housing Anywhere, 2023; Pandiyan, 2023). However, in Spain, the yearly average salary is €24,009 while a Spanish nurse makes €39,458 a year (The Local, 2022; Erieri, 2022). So, this finding seems to agree with the statement that a higher household salary increases loss aversion. Still, the contrasting literature makes it difficult to determine if the difference in survey response rate is a result of the country its characteristics, or if it is due to coincidence which may be more likely as the experiment has only run once.

The result of the survey response rate regarding loss aversion did not show a correlation between the BE principle, response rate and country, so no further analysis will be done for this BE principle.

Overall Impression

When comparing the studies that used methods to experiment with the same BE principles of this thesis, but for other goals and in different industries, it appears that the scope that was used in this research was too narrow. Studies that had been successful in using BE principles to achieve their goals used multiple aspects that could have had an impact on the effectiveness of the BE principles. For example, it may have been better if the BE principles were able to influence the survey participants in person to have a greater effect. However, this was not possible in this research, because the researcher had to be pragmatic with the available resources and time. Additionally, constant or repeated exposure to the BE principle would have increased the intensity of its effect and potentially be beneficial to achieve the goal of the desired behaviour. The characteristics of the participants should also be taken into consideration well, such as cognitive load and the intrusion of the research having a negative effect on the results. Perhaps, cultural aspects may have an impact on how effective BE principles are, as some BE principles are more accepted in certain cultures than others. Finally, some additional design decisions could have further improved the effectiveness of the BE principles.

Furthermore, across all these fields there are few articles published about how BE influences survey response rate specifically. While surveys are often used to question participants on how certain scenarios relating to BE principles would affect behaviour, the majority of these studies do not focus on improving the survey response rate. In the studies that did attempt to use BE as a method to boost survey response rate, such as the study by Baruch & Holtom (2008), the result of using incentives was found not to be related to survey response rate and the use of reminders was found to have a lower response rate. Another study where social incentives were used as an attempt to boost the survey response rate was also not successful in improving the overall survey response rate (Warwick, et al., 2019).

Following these results, the question if BE affects survey response rate positively may be answered. From the results of this research and the results of other studies testing how BE principles can improve survey response rate, it suggests that the usage of BE is not appropriate for this goal. Although some flaws in this research may have influenced the result, BE is not an appropriate method to improve the survey response rate. Even more so when in-person interactions are unable to be executed, such as an in-app survey such as the one used in this research.

4.5.2. Benefits of Behavioural Economics

The use of BE has multiple benefits that became clear throughout the research. These benefits will be discussed in this section.

Informed decision-making – The result that comes out of experimenting with BE is easy to interpret and supports informed decision-making. The result provides data and insights in such a way that this can be evaluated on how changes made to the survey impact performance. This result proves which change works and which does not, which allows for making decisions about how the survey should be designed. So, the result allows for the identification of these areas for improvement and more informed decisions.

Ethical considerations – For the implementation of BE, most guidelines highlight ethical considerations. After the ethical concerns that were raised with the Milgram experiment (1963), conducting ethical research has been pushed. Therefore, it helps the researcher with considering potential ethical concerns and adjusting the experiment in such a way that participants in the experiment are not harmed. In this research, breaches of the privacy policy were discussed and adjusted. Following this, the users of the application were informed and the research was done anonymously. Additionally, the researcher kept the design user-friendly by giving the participants freedom of choice by clicking away from the survey. This ensured the research was ethically sound.

Assumption proof – Instead of assuming that the desired behaviour change is achieved by its mere implementation according to theory, BE allows researchers to experiment with real people without being intrusive. This way, human behaviour can be tested and realistically represented as it allows researchers to understand how humans behave in unideal real-world situations and how it impacts decision-making.

4.5.3. Challenges of Behavioural Economics

Besides benefits, there were also multiple challenges that were faced during the research when using BE. These challenges will be discussed in this section.

Consistency – When conducting the literature review for this research, many different opinions on the usage of BE for changing desired behaviour were found. While some studies tend to praise the potential of BE following positive results, similar studies have also been found where BE failed to change the behaviour of experiment participants to the desired behaviour. Similarly to this research, the BE principles of this research had been chosen based on their potential, but as a result were not effective. The inconsistency across study results makes it very challenging to assess reliable BE theories and consequently decide on BE principles to use, which in this research has led to the used BE principles in this study being ineffective.

Influence of external factors – Human behaviour is sensitive and is heavily influenced by external factors. These external factors can be temporary (emotional) or long-term (culture). Therefore, it becomes difficult to predict human behaviour when someone is confronted with making a decision. In this research where the culture of the participants differs, even slightly, generalizing the BE interventions in such a way for it to be effective across the entire sample is very challenging. Additionally, the high work pressure that is experienced by the participants has an influence on their emotional and cognitive aspects which leads to changes in expected behaviour. These external factors are difficult to impossible to control and may have led to negatively impacting the results.

Practical considerations – As mentioned in the previous challenge, it proves difficult to translate theory into practical BE interventions for a multitude of reasons. One reason is that ethical considerations need to be taken into account and may make it more difficult to apply the intervention as intended. Although this did not entirely skew the design for this research, it did cause an inevitable delay. Another is that resource and time constraints made it difficult to design interventions in such a way as other studies had done, such as more physical interventions.

4.5.4. Limitations

This research faced multiple limitations in its execution. These limitations will be discussed in this section.

Limited scope – This research is able to cover only 3 of the 52 principles of the MINDSPACE framework in the available time. So, although a conclusion can be made about the principles that were tested, these are not able to represent BE as a whole which would lead to selection bias where only certain principles are taken into account. Furthermore, the scope of the research may have been too narrow due to time and resource constraints. As a result, the BE interventions were not able to be designed in such a way that other studies have done with positive results.

Bias – The research methodology that was used for this study was action research. However, action research is vulnerable to bias. Some of these biases are but are not limited to contextual bias, ethical bias, selection bias, and confirmation bias (Simundic, 2013; McKay & Marshall,

2002; Hallihan & Shu, 2013). As mentioned, there is a likelihood of contextual bias having affected the results of this research. The reasoning behind this is how the emotional state and culture of the participants could have influenced their decision-making when they had received the survey. However, ethical bias is not a concern in this research as the choice made that could affect the participants have been carefully considered. Furthermore, selection bias is not a concern in this study as prior to the research it had been defined what the target population would be. Still, this specific target population may prove a generalization of the result to be difficult. Finally, confirmation bias is not likely to be a concern in this research either, as the quantitative results of the study are objectively compared against each other to make a conclusion. So, the limitations regarding bias in this study may be due to the contextual situation of the participants which are difficult to control as mentioned in the challenges of BE.

Literature gap – Upon the start of the research, there was little work available on how BE has been tested in a survey setting. So, there is no reliable reference point for the result of this research. Consequently, no further conclusion can be made on how well or how poorly the result of this research was in comparison to the results of other experiments other than the conclusion that BE did not perform well in this research.

Repetition – Due to time constraints, it was not possible to repeat the experiment. Action research is an iterative process and behavioural studies highly recommend repeating experiments to average out the results. However, this is not possible and may have impacted the results.

4.5.5. Areas of Improvement

From the experience of managing the research, conducting the experiment and the results that were received, some areas of improvement can be identified.

As for the design of the experiment, a broader scope could have been taken to improve the intensity of the BE principles that were used to leverage their potential in changing behaviour. However, as mentioned, this was not possible with the time and resources available for this research.

Another area of improvement is the repetition of the experiment. As mentioned as part of the limitations, the experiment was only run once because of time constraints. Because of this, the results did not have another result to compare to and could not be averaged out. Additionally, due to this, the researcher was unable to account for any possible mistakes made during the data collection phase.

4.5.6. Future Research

During the research, it became clear that there was a literature gap between the effectivity of BE on survey response rate. However, with the result of this research, it does not appear that BE is an appropriate tool to improve the survey response rate. It is likely that this topic suffered from publication bias as there are so few studies available about BE being effective to boost survey response rate. Those studies that are available often use BE for other behavioural changes in other fields. So, it is recommended against research in the same area as this research as no significant value can be added.

5. Conclusion

With feedback being an important aspect of evaluating and improving a service or product, even more so in highly competitive industries, receiving a high survey response rate on CX feedback surveys is desired. The goal of this research was to determine if BE is an effective method for increasing survey response rate by experimenting with three BE principles through A/B testing. First, the concept of BE was explored and three BE principles to be tested were chosen. Second, the three BE principles were applied to the survey designs. Before testing the surveys, technical decisions were made on how the surveys would be distributed among the sample population and when these surveys would be showing up. Also, the decisions were made on what data would be collected and where this would be stored. Finally, the surveys were implemented in the application and sent out to the users to start with gathering the data necessary to calculate the response rate of each BE principle as well as the control group that did not receive a survey with a BE principle implemented into it. This research contributes to evaluating BE in the context of in-app surveys and provides a recommendation for the usage of the three BE principles in this specific context to assist in improving the survey response rate for organisations.

The main findings of this research are that the three tested BE principles (authority bias, social influence, and loss aversion) received a lower survey response rate in the experiment than the control group which received the survey without BE. The result of the chi-squared test indicated that the survey version did have an effect on the survey response rate. As the non-BE version of the survey received the highest survey response rate, the three BE principles are not appropriate to be used in the in-app survey context as their effect. Due to the limitations in this research, the BE principles may have not had the intended effect on behaviour change as certain aspects of BE interventions that were used in other BE studies were not able to be implemented. Other aspects had not been taken into account and may have had an influence on the results. The broad variety of aspects found in BE interventions is shown by the high heterogeneity in the result of the meta-analysis. The meta-analysis used multiple studies that used incentives as their BE intervention, yet yielded different results. Even across the studies where different BE principles were used as the intervention, different results were achieved. Although the overall effect of the combined studies slightly showed favour for the BE interventions, the confidence intervals cross the null hypothesis line, causing the result to be not statistically significant. When removing the studies that cause the high heterogeneity, the overall effect becomes statistically significant and still is slightly in favour of the BE interventions.

Following this narrative and as the BE interventions in this study were not effective, it is not recommended for researchers to continue researching the effectiveness of these BE principles on survey response rate. Furthermore, as the results of using the same BE principles still yield such different results, it is recommended against further researching BE as a whole. For organisations, it is also recommended against applying BE in surveys to attempt to boost the survey response rate. Not only does this require more resources to execute, but it is also not a reliable tool to boost the survey response rate with.

In conclusion, the BE principles in this thesis are not effective to improve the survey response rate when applied to in-app surveys. Aspects that were used in other studies to make BE interventions effective are too difficult to be used in the in-app survey context, such as physical nudges. Therefore, BE was applied too narrowly, but this is the result of the limitations of an in-app survey. Furthermore, it is important to be cautious about the interpretation of the meta-analysis. Despite the result being slightly in favour of BE interventions, the main result was statistically insignificant. If this narrative is followed, it is recommended against using authority bias, social influence, and loss aversion to boost the survey response rate. The result of the chi-squared test supports this by indicating that the control group performed better than the treatment group which was not due to chance. So, it is recommended against applying these three BE principles in an attempt to boost the survey response rate. Additionally, because the results of the studies from the meta-analysis had such high variability and the studies were heterogeneous, it is not recommended to continue research on BE as a method to boost the survey response rate. The practical contribution of this research is that it creates a foundation for other researchers. Additionally, it also recommends against further research on how BE can be tested in the context of in-app surveys to improve the survey response rate. The academic contribution of this research is to provide further insight into the application of BE to in-app surveys and its effectiveness on survey response rate and provide additional knowledge to fill the literature gap.

6. References

Abbott, L. (1956). Quality and Competition. *American Economic Association*, 171-174.

Alcott, H. (2011). Social norms and energy conservation. *Journal of Public Economics*, 1082-1095.

Alessi, E. J., & Martin, J. I. (2010). Conducting an Internet-based Survey: Benefits, Pitfalls, and Lessons Learned. *Social Work Research*, 122-128.

Anderson, E., Fornell, C., & Lehmann, D. (1994). Customer Satisfaction, Market Share, and Profitability: Findings from Sweden. *Journal of Marketing*, 53-66.

Anderson, E., Fornell, C., & Rust, R. (1997). Customer Satisfaction, Productivity, and Profitability: Differences Between Goods and Services. *Marketing Science*, 97-181.

Askariazad, M., & Babakhani, N. (2015). An application of European Customer Satisfaction Index (ECSI) in business to business (B2B) context. *Journal of Business & Industrial Marketing*, 17-31.

Awwad, M. (2012). An application of the American Customer Satisfaction Index (ACSI) in the Jordanian mobile phone sector. *The TQM Journal*, 529-540.

Baethge, A., & Rigotti, T. (2013). Interruptions to workflow: Their relationship with irritation and satisfaction with performance, and the mediating roles of time pressure and mental demands. *An International Journal of Work, Health & Organisations*, 43-63.

Bain. (2015, April 8). *Are you experienced?* (Bain) Retrieved January 12, 2022, from https://www.bain.com/insights/are-you-experienced-infographic/

Bain. (2018). *NEW BAIN CERTIFIED NPS BENCHMARKS: COMPANIES CUSTOMERS LOVE.* Retrieved from Bain: https://media.bain.com/Images/BAIN_DIGEST_New_Bain_Certified_NPS_Benchmarks.pdf

Bakshy, E., Eckles, D., Yan, R., & Rosenn, I. (2012). Social Influence in Social Advertising: Evidence from Field Experiments. *Proceedings of the 13th ACM Conference on Electronic Commerce* (pp. 146-161). New York: ACM.

Baruch, Y. (1999). Response Rate in Academic Studies-A Comparative Analysis. *Human Relations*.

Baruch, Y., & Holtom, B. (2008). Survey response rate levels and trends in organizational research. *Human Relations*.

BBC News. (2019, July 1). *How rude service inspired a multi-million euro firm*. Retrieved from BBC: https://www.bbc.com/news/business-48760522

Bećirović, D., Suhonjić, A., & Stanić, M. (2022). USING LOSS AVERSION AND FRAMING TO NUDGE. *Journal of Contemporary Management Issues*, 5-17.

Becker, L., & Jaakkola, E. (2020). Customer experience: fundamental premises and implications for research. *Journal of the Academy of Marketing Science*, 630-648.

BI Team. (2014, April 11). *EAST: Four Simple Ways to Apply Behavioural Insights.* Retrieved from BI: https://www.bi.team/wp-content/uploads/2015/07/BIT-Publication-EAST_FA_WEB.pdf

Biskirk, T. D., & Andres, C. (2012). Smart Surveys for Smart Phones: Exploring Various Approaches for Conducting Online Mobile Surveys via Smartphones. *Survey Practice*.

Blumenberg, C., & Barros, A. (2018). Response rate differences between web and alternative data collection methods for public health research: a systematic review of the literature. *International Journal of Public Health*, 765-773.

Bojinov, I., Saint-Jacques, G., & Tingley, M. (2020). Avoid the Pitfalls of A/B Testing. *Harvard Business Review*.

Bond, R. F. (2012). A 61-million-person experiment in social influence and political mobilization. *Nature*, 295-298.

Burger, J. (2009). Replicating Milgram: Would people still obey today? *American Psychologist*, 916-925.

Buttle, F. (1994). SERVQUAL: review, critique, research agenda. *European Journal of Marketing*, 8-32.

Byrd, T. A., & Turner, D. E. (2000). Construct, Measuring the Flexibility of Information Technology Infrastructure: Exploratory Analysis of a. *Journal of Management Information Systems, 17*(1), 167-208.

Cano, S., Rusu, C., & Quiñones, D. (2020). Evaluation of Customer eXperience and Behaviour: A Literature Review. *Internation Conference on Human-Computer Interation.*

Capurro, D., & Velloso, E. (2021). Dark Patterns, Electronic Medical Records, and the Opioid Epidemic. *arXiv preprint*.

Carbone, L. P., & Haeckel, S. H. (1994). Engineering Customer Experiences. *American Marketing Association*.

Casula, M., Rangarajan, N., & Shields, P. (2021). The potential of working hypotheses for deductive exploratory research. *Quality & Quantity*, 1703-1725.

Chakraborty, A. (2019, Januari 1). *Authority Bias: Too much obedience might make you do stupid things.Ch*. Retrieved from Medium: https://medium.com/@coffeeandjunk/cognitive-bias-authority-bias-5e0e10ee677c

Cialdini, R. B. (1998). Social influence: Social norms, conformity and compliance. *The handbook of social psychology*, 151-192.

Clear, J. (2018). *Atomic Habits.* Random House Business.

Clearly Cultural. (2022). *POWER DISTANCE INDEX*. Retrieved from Clearly Cultural: https://clearlycultural.com/geert-hofstede-cultural-dimensions/power-distance-index/

Cohn, A., & Maréchal, M. (2016). Priming in economics. *Current Opinion in Psychology*, 17-21.

Cole, R. E. (2002). From continuous improvement to continuous innovation. *Total Quality Management, 13*(2), 1051-1056.

Connors, E., Krupnikov, Yanna, & Ryan, J. (2019). HOW TRANSPARENCY AFFECTS SURVEY RESPONSES. *Public Opinion Quarterly*, 185-209.

Cook, D., Wittich, C., Daniels, W., West, C., Harris, A., & Beebe, T. (2016). Incentive and Reminder Strategies to Improve Response Rate for Internet-Based Physician Surveys:A Randomized Experiment. *JOURNAL OF MEDICAL INTERNET RESEARCH*.

Costa, D., & Kahn, M. (2013). Energy Conservation "Nudges" and Environmentalist Ideology: Evidence from a Randomized Residential Electricity Field Experiment . *Journal of the European Economic Association*, 680-702.

Cuttica, P. (2016, April 18). *Quantify Your Customer Care Efforts With Twitter Customer Feedback in Sprout Social*. Retrieved from Sprout Social: https://sproutsocial.com/insights/twitter-customer-care/

Daikeler, J., Silber, H., & Bosnjak, M. (2021). A Meta-Analysis of How Country-Level Factors Affect Web Survey Response Rates. *International Journal of Market Research*, 1-28.

Davison, R., Martinsons, M., & Kock, N. (2004). Principles of canonical action research. *Info Systems*, 65-86.

Davison, R., Martinsons, M., & Ou, C. (2012). The Roles of Theory in Canonical Action Research. *MIS Quarterly*, 763-786.

De Haan, E., Verhoef, P., & Wiesel, T. (2015). The predictive ability of different customer feedback metrics for retention. *Inter. J. of Research in Marketing*, 195-206.

De Keyser, A., Lemon, K. N., Klaus, P., & Keiningham, T. L. (2015). A Framework for Understanding and Managing the Customer Experience. *Marketing Science Institute Working Paper Series*, 1-47.

Deng, W., Yeh, M., & Sung, M. (2013). A customer satisfaction index model for international tourist hotels: Integrating consumption emotions into the American Customer Satisfaction Index. *International Journal of Hospitality Management*, 133-140.

Deutsch, M., & Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgment. *Journal of Abnormal and Social Psychology*, 629-636.

Devine, J., & Gilson, K. (2010, February 1). *Using behavioral science to improve the customer experience*. (McKinsey) Retrieved January 23, 2022, from https://www.mckinsey.com/business-functions/operations/our-insights/using-behavioral-science-to-improve-the-customer-experience

Diebner, R., Malfara, D., Neher, K., Thompson, M., & Vancauwenberge, M. (2021, February 24). Prediction: The Future of CX. *McKinsey Quarterly*. Retrieved from McKinsey: https://www.mckinsey.com/business-functions/marketing-and-sales/our-insights/prediction-the-future-of-cx

Dijksterhuis, A., & Bargh, J. (2001). The perception–behavior expressway: Automatic effects of social perception on social behavior. *Advances in experimental social psychology*, 1-40.

Dixon, M., Freeman, K., & Toman, N. (2010). Stop Trying to Delight Your Customers. *Harvard Business Review*.

Dolan, P., Hallsworth, M., Halpern, D., D., K., Metcalfe, R., & Vlaev, I. (2011). Influencing behaviour: The mindspace way. *Journal of Economic Psychology*, 264-277.

Dolan, P., Hallsworth, M., Halpern, D., King, D., & Vlaev, I. (2010). *Going with the grain: influencing behaviour through public policy.* BI Team.

Donorregister. (2020). *Voorlichtingsmateriaal*. Retrieved from Donorregister: https://www.donorregister.nl/voorlichtingsmateriaal/english

Donorregister. (2022). *Over het donorregister*. Retrieved from Donorregister: https://www.donorregister.nl/over-het-donorregister/cijfers

Eberhardt, W., Brüggen, E., Post, T., & Hoet, C. (2021). Engagement behavior and financial well-being: The effect of message framing in online pension communication. *International Journal of Research in Marketing*, 448-471.

Edwards, P., Roberts, I., Clarke, M., DiGuiseppi, C., Wentz, R., Kwan, I., . . . Pratap, S. (2009). Methods to increase response to postal and electronic. *Edwards PJ, Roberts I, Clarke MJ, DiGuiseppi C, Wentz R, Kwan I, Cooper R, Felix LM, Pratap S*, 1-394.

Eklöf, J. (2000). *European Customer Satisfaction Index Pan-European Telecommunication Sector Report Based on the Pilot Studies.* European Organization for Quality and European Foundation for Quality Management.

Elfering, A., Grebner, S., & Ebener, C. (2015). Workflow interruptions, cognitive failure and near-accidents in health care. *Psychology, Health & Medicine*, 138-147.

Emdad, A. F., & Bhatt, G. D. (2009). An Empirical Examination of the Relationship between Information Technology Infastructure, Customer Focus and Business Advantages. *Americas Conference on Information Systems.*

Erieri. (2022). *Registered Nurse Salary in Spain*. Retrieved from Erieri: https://www.erieri.com/salary/job/registered-nurse/spain

Eugenia Cadenas Sáez, M. (2020). Behavioural economics for investor protection. *CNVM*, 1-35.

Fabijan, A., Dmitriev, P., Olsson, H., & Bosch, J. (2016). Online Controlled Experimentation at Scale: An Empirical Survey on the Current State of A/B Testing. *2016 International Conference on Computing, Communication and Automation.* IEEE.

Fan, W., & Yan, Z. (2010). Factors affecting response rates of the web survey: A systematic review. *Computers in Human Behaviour*, 132-139.

Fincham, J. (2008). Response Rates and Responsiveness for Surveys, Standards, and the Journal. *American Journal of Pharmaceutical Education*.

Fornell, C. (1992). A National Customer Satisfaction Barometer: The Swedish Experience. *Journal of Marketing*, 6-21.

Fornell, C., Johnson, M., Anders, E., Cha, J., & Bryant, B. (1996). The American Customer Satisfaction Index: Nature, Purpose, and Findings. *American Marketing Association*, 7-18.

Frey, B. (2018). *The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation.* SAGE Publications, Inc.

Fryer, R., Levitt, S., List, J., & Sadoff, S. (2012). Enhancing the Efficacy of Teacher Incentives through Loss Aversion: A Field Experiment. *NATIONAL BUREAU OF ECONOMIC RESEARCH*.

Gächter, S., Johnson, E., & Herrmann, A. (2007). Individual-Level loss aversion in riskless and risky choices. *IZA Discussion Paper*.

Gallo, A. (2017). A Refresher on A/B Testing. *Harvard Business Review*.

Gearson, M. (2019, October 16). *Cognitive biases: loss aversion*. Retrieved from UX Design: https://uxdesign.cc/cognitive-biases-loss-aversion-925149360f46

Gentile, C., Spiller, N., & Noci, G. (2007). How to Sustain the Customer Experience: An Overview of Experience Components that Co-create Value With the Customer. *European Management Journal*, 395-410.

Gregoire, C., & Kaufman, S. (2016, February 22). *Bias Against Creativity*. Retrieved from Behavioural Scientist: https://behavioralscientist.org/the-bias-against-creativity/

Hallek, M., Ockenfels, A., & Wiesen, D. (2022). Behavioral Economics Interventions to Improve Medical Decision-Making. *Dtsch Arztebl Int*, 633-639.

Hallihan, G., & Shu, L. (2013). Considering Confirmation Bias in Design and Design. *ournal of Integrated Design & Process Science*, 19-35.

Harris, M. (2010). Invited Commentary: Evaluating Epidemiologic Research Methods—The Importance of Response Rate Calculation . *American Journal of Epidemiology*, 645-647.

Hathaway, C., Chavez, M., Kadono, M., Ketcher, D., Rollinson, D., Siegel, E., . . . Gonzalez, B. (2021). Improving Electronic Survey Response Rates Among Cancer Center Patients During the COVID-19 Pandemic: Mixed Methods Pilot Study. *JMIR CANCER*.

Hirschman, A. (1970). *Exit, Voice, and Loyalty: Responses to Decline in Firms, Organizations, and States.* Harvard University Press.

Hofling, C., Brotzman, E., Dalrymple, S., Graves, N., & Pierce, C. (1966). An experimental study in nurse-physician relationships. *The Journal of Nervous and Mental Disease*, 171-180.

Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors and institutions across nations.* Sage Publications.

Holtom, B., Baruch, Y., Aguinis, H., & Ballinger, G. (2022). Survey response rates: Trends and a validity assessment framework. *Human Relations*, 1560-1584.

Housing Anywhere. (2023, January 25). *Your Guide to Average Salaries in Germany (2022–2023).* Retrieved from Housing Anywhere: https://housinganywhere.com/Germany/average-salaries-in-germany-2021

Hoyer, W. D., Kroschke, M., Schmitt, B., Kraume, K., & Shankar, V. (2020). Transforming the Customer Experience Through New Technologies. *Journal of Interactive Marketing, 51*, 57-71.

Jin, L. (2011). Improving Response Rates in Web Surveys with Default Setting: The Effects of Default on Web Survey Participation and Permission. *International Journal of Market Research*, 75-94.

Kahneman, D. (2011). *Thinking, Fast and Slow.* Farrar, Straus and Giroux.

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 263-292.

Kahneman, D., & Tversky, A. (1992). Advances in Prospect Theory: Cumulative Representation of Uncertainty . *Journal of Risk and Uncertainty*, 297-323.

Kemmis, S., McTaggart, R., & Nixon, R. (2013). *The Action Research Planner: Doing Critical Participatory Action Research.* Springer Science & Business Media.

Kenton, W. (2020, September 28). *Behavioral Economics*. Retrieved from Investopedia: https://www.investopedia.com/terms/b/behavioraleconomics.asp

Kivetz, R., Urminsky, O., & Zheng, Y. (2006). The Goal-Gradient Hypothesis Resurrected: Purchase Acceleration, Illusionary Goal Progress, and Customer Retention. *Journal of Marketing Research*, 39-58.

Kohavi, R., & Longbotham, R. (2017). Online Controlled Experiments and A/B Testing. *Encyclopedia of Machine Learning and Data Mining*.

Kohavi, R., & Thomke, S. (2017, September). *The Surprising Power of Online Experiments*. Retrieved from Harvard Business Review: https://hbr.org/2017/09/the-surprising-power-of-online-experiments

Kviz, F. (1977). Toward a Standard Definition of Response Rate. *Oxford Journals*, 265-267.

Laibson, D., & List, J. (2015). Principles of (Behavioral) Economics. *American Economic Review*, 385-390.

Larsen, N., Stallrich, J., Sengupta, S., & Deng, A. (2022). Statistical Challenges in Online Controlled Experiments: A Review of A/B Testing Methodology. *ArXiv*.

Lashbrook, J. (2000). Fitting in: exploring the emotional dimension of adolescent peer pressure. *Adolescence*, 747-757.

Lemon, K. N., & Verhoef, P. C. (2016). Understanding Customer Experience Throughout the Customer Journey. *Journal of Marketing*, 69-96.

Lindgren, E., Markstedt, E., Martinsson, J., & Andreasson, M. (2018). Invitation Timing and Participation Rates in Online Panels: Findings From Two Survey Experiments. *Social Science Computer Review*, 1-20.

Lucas, H. C., & Olson, M. (1994). FLEXIBILITY, THE IMPACT OF INFORMATION TECHNOLOGY ON ORGANIZATIONAL. *Journal of Organizational Computing, 4*(2), 144-176.

Madrian, B. C. (2014). APPLYING INSIGHTS FROM BEHAVIORAL ECONOMICS TO POLICY DESIGN. *Annual Review of Economics*.

Maymone, M., Venkatesh, S., Secemsky, E., Reddy, K., & Vashi, N. (2018). Research Techniques Made Simple: Web-Based Survey Research in Dermatology: Conduct and Applications. *Journal of Investigative Dermatology*, 1456.

McGuinness, L. A., & Higging, J. P. (2021). Risk-of-bias VISualization (robvis): An R package and Shiny web app for visualizing risk-of-bias assessments. *Res Synth Methods*, 55-61.

McKay, J., & Marshall, P. (2002). Action Research: A Guide to Process and Procedure. *European Conference on Research Methods in Business and Management.*

Meija, M. (2021). Theory-Driven or Theory-Informed? A Review of Behavioural Economics in Design. *The Design Journal*, 1-30.

Meyer, C., & Schwager, A. (2007). Understanding Customer Experience. *Harvard Business Review*, 116.

Milgram, S. (1963). Behavioral Study of Obedience. *The Journal of Abnormal and Social Psychology*, 371-378.

Milgram, S. (1963). Behavioral Study of Obedience. *Journal of Abnormal and Social Psychology*, 371-378.

Millis, G. (2017). *Action Research A Guide For The Teacher Researcher.* Pearson.

Mueller, K., Straatmann, T., Gattrup, K., & Jochum, M. (2014). Effects of Personalized Versus Generic Implementation of an Intra-Organizational Online Survey on Psychological Anonymity and Response Behavior: A Field Experiment. *Journal of Business Pyschology*, 169-181.

Mullainathan, S., & Thaler, R. H. (2000). Behavioral Economics. *International Encyclopedia of the Social and Behavioral Sciences*.

Nattinger, A., Pezzin, L., Sparapani, R., Neuner, J., King, T., & Laud, P. (2010). Heightened Attention to Medical Privacy: Challenges for Unbiased Sample Recruitment and a Possible Solution. *American Journal of Epidemiology*, 637-644.

Nazi, K. (2010). Veterans' voices: use of the American Customer Satisfaction Index (ACSI) Survey to identify My HealtheVet personal health record users' characteristics, needs, and preferences. *Journal of the American Medical Informatics Association*, 203-211.

Ness, L. (2008). Exploring Information Technology Flexibility and Increased Complexity: Is Management Prepared? *The International Journal of Applied Management and Technology, 6*(3).

Nichols, E., Konya, S., Horwitz, R., & Raim, A. (2019). *2020 Census Research and Testing Report: The Effect of the Mail Delivery Date on Survey Login Rates and Helpline Call Rates.* Washington: U.S. Census Bureau.

Nolan, J. M., Schultz, P. W., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2011). Normative social influence is underdetected. *Personality and Social Psychology Bulletin*, 1513-1523.

Novemsky, N., & Kahneman, D. (2005). The Boundaries of Loss Aversion. *Journal of Marketing Research*, 119-129.

OECD. (2019, June 18). *Tools and Ethics for Applied Behavioural Insights: The BASIC Toolkit.* Retrieved from OECD: https://www.oecd.org/gov/regulatory-policy/BASIC-Toolkit-web.pdf

Olenski, S. (2017, July 20). *How Brands Should Use Celebrities For Endorsements*. Retrieved from Forbes: https://www.forbes.com/sites/steveolenski/2016/07/20/how-brands-should-use-celebrities-for-endorsements/?sh=1282fc655936

O'Loughlin, C., & Coenders, G. (2002). Application of the European Customer Satisfaction Index to Postal Services. Structural Equation Models versus Partial Least Squares. *Research Papers in Econmics*.

O'Reilly-Shah, V. N. (2017). Factors influencing healthcare provider respondent fatigue answering a globally administered in-app survey. *PeerJ Life & Environment*.

Pachler, D., Kuonath, A., Specht, J., Kennecke, S., Agthe, M., & Frey, D. (2018). Workflow interruptions and employee work outcomes: The moderating role of polychronicity. *J Occup Health Psychol.*, 417-427.

Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, C, T., & Mulrow, C. D. (2021). *PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews.* BMJ.

Pandiyan, D. (2023, March 9). *Nurse Salary In Germany*. Retrieved from Kanan: https://www.kanan.co/blog/nurse-salary-in-germany/

Payne, J., Samper, A., Bettman, J. R., & Luce, M. F. (2008). Boundary conditions on unconscious thought in complex decision making. *Psychological Science*, 1118-1123.

Pine II, B. J., & Gilmore, J. H. (1998). Welcome to the Experience Economy. *Harvard Business Review*.

Polonioli, A., Ghioni, R., Greco, C., Juneja, P., Tagliabue, J., Watson, D., & Floridi, L. (2022). The Ethics of Online Controlled Experiments (A/B testing). *SSRN*, 1-30.

Qualtrics. (2021, November 23). *What is CSAT (customer satisfaction score)?* Retrieved from Qualtrics: https://www.qualtrics.com/uk/experience-management/customer/what-is-csat/?rid=ip&prevsite=en&newsite=uk&geo=NL&geomatch=uk

Qualtrics. (2022, April 15). *How to measure customer satisfaction: 4 key metrics*. Retrieved from Qualtrics: https://www.qualtrics.com/experience-management/customer/measure-customer-satisfaction/

Qualtrics. (2022, February 17). *How to write good survey questions*. Retrieved from Qualtrics: https://www.qualtrics.com/uk/experience-management/research/survey-questions/?rid=ip&prevsite=en&newsite=uk&geo=NL&geomatch=uk

Rackoff, G., Monocella, L., Fowler, L., Vazquez, M., Shah, J., Fitzsimmons-Craft, E., . . . Newman, M. (2023). Using Social Influence Strategies to Improve Rates of Online Mental Health Survey Participation: Results from Two Experiments. *Journal of Behavioral and Cognitive Therapy*, 1-30.

Reichheld, F. (2003, December). *The One Number You Need to Grow*. Retrieved from Harvard Business Review: https://hbr.org/2003/12/the-one-number-you-need-to-grow

Reichheld, F. (2006). *The Ultimate Question: Driving Good Profits and True Growth.* Harvard Business School Press.

Reisch, L., & Sunstein, C. (2023). Do Europeans like nudges? *Judgement and Decision Making*.

Rogelberg, S., Luong, A., Sederburg, M., & Cristol, D. (2000). Employee attitude surveys: examining the attitudes of noncompliant employees. *The Journal of Applied Psychology*.

Rudenko, L. A., Smiianov, V. A., & Smiianova, O. I. (2020). BASIC PRINCIPLES OF BEHAVIORAL ECONOMICS AND PROSPECTS FOR THEIR APPLICATION IN THE PUBLIC HEALTH SYSTEM. *Wiadomości Lekarskie*, 2026-2030.

Sánchez-Fernández, J., Muñoz-Leiva, F., & Montoro-Ríos, F. J. (2012). Improving retention rate and response quality in Web-based surveys. *Computers in Human Behavior*, 507-514.

Saunders, M., Lewis, P., & Thornhill, A. (2019). Research Methods for Business Students. In M. Saunders, P. Lewis, & A. Thornhill, *Research Methods for Business Students* (pp. 128-171). Pearson Education.

Schlüter, M., Baeza, A., Dressler, G., Frank, K., Groeneveld, J., & Jager, W. (2017). A framework for mapping and comparing behavioural theories in models of social-ecological systems. *Ecological Economics*, 21-35.

Schmatz, R., Wolf, G., & Landmann, M. (2015). Students as customers - The Net Promoter Score as a measure of satisfaction and loyalty in higher education. *From here to there: Positioning higher education institutions.* Krems.

Schwager, A., & Meyer, C. (2007). Understanding Customer Experience. *Harvard Business Review*.

Scribbr. (2022, May 31). *Chi-Square (X²) Table | Examples & Downloadable Table*. Retrieved from Scribbr: https://www.scribbr.com/statistics/chi-square-distribution-table/

Seshadri, S., & Broekmier, G. (2022). Small Business Executives' Online Survey Response Intentions: The Effects of Incentives and Survey Length. *Small Business Institute Journal*, 1-9.

Shang, J., Basil, D., & Wymer, W. (2010). Using social marketing to enhance hotel reuse programs. *Journal of Business Research*, 166-172.

Shaw, C., & Ivens, J. (2002). *Building Great Customer Experiences.* London: Palgrave Macmillan.

Simundic, A.-M. (2013). Bias in Research. *Biochemia Medica*, 12-15.

Stanley, M., Roycroft, J., Amaya, A., Dever, J. A., & Srivastav, A. (2020). The Effectiveness of Incentives on Completion Rates, Data Quality, and Nonresponse Bias in a Probability-based Internet Panel Survey. *Field Methods*, 159-170.

Sun, K.-A., & Kim, D.-Y. (2013). Does customer satisfaction increase firm performance? An application of American Customer Satisfaction Index (ACSI). *International Journal of Hospitality Management*, 68-77.

Taylor, S., & Thompson, S. (1982). Stalking the elusive "vividness" effect. *Psychological Review*, 155-181.

Terblanche, N. (2006). An application of the American customer satisfaction index (ACSI) in the South African motor vehicle industry. *South African Journal of Business Management*, 29-38.

Thaler, R. (2016). Behavioral economics: Past, present, and future. *American Economic Review*, 1577-1600.

Thaler, R. H., & Sunstein, C. R. (2009). *Nudge.* Penguin Putnam Inc.

The ASCI. (2022, March 9). *Our Industries*. Retrieved from The ASCI: https://www.theacsi.org/our-industries/

The Decision Lab. (2018). *MINDSPACE Framework*. Retrieved from The Decision Lab: https://thedecisionlab.com/reference-guide/neuroscience/mindspace-framework

The Local. (2022, October 11). *What is Considered a Decent Salary in Spain?* Retrieved from The Local: https://www.thelocal.es/20221011/what-is-considered-a-decent-salary-in-spain

Tiwari, P. (2021). Effect of Media on the Behaviour of Investors and Stocks. *Turkish Online Journal of Qualitative Inquiry*, 1667-1673.

Twitter. (2022). *Customer feedback*. Retrieved from Twitter: https://business.twitter.com/en/help/campaign-measurement-and-analytics/customer-feedback.html

van Haaften, R. (2017). *Customer satisfaction models*. Retrieved from van Haaften: https://www.van-haaften.nl/customer-satisfaction/customer-satisfaction-models/61-the-european-customer-satisfaction-index

Van Ryzin, G., Muzzio, D., Immerwahr, S., Gulick, L., & Martinez, E. (2004). Drivers and Consequences of Citizen Satisfaction: An Application of the American Customer Satisfaction Index Model to New York City. *Public Administration Review*, 331-341.

Vangeest, J., & Johnson, T. (2011). Surveying Nurses Identifying Strategies to Improve Participation. *Evaluation & the Health Professions* , 487-511.

Verhoef, P. C., Roggeveen, A. L., Parasuraman, A. P., & Tsiros, M. (2009). Customer Experience Creation: Determinants, Dynamics and Management Strategies. *Journal of Retailing*, 31-41.

Wang, M., Rieger, M., & Hens, T. (2017). The impact of culture on loss aversion. *Journal of Behavioral Decision Making*, 270-281.

Warwick, H., Hutyra, C., Politzer, C., Francis, A., Risoli, T., Green, C., . . . Mather, R. (2019). Small Social Incentives Did Not Improve the Survey Response Rate of Patients Who Underwent Orthopaedic Surgery: A Randomized Trial. *Clin Orthop Relat Res*, 1648-1656.

Warwick, H., Hutyra, C., Politzer, C., Francis, A., Risoli, T., Green, C., . . . Mather, R. (2019). Small Social Incentives Did Not Improve the Survey Response Rate of Patients Who Underwent Orthopaedic Surgery: A Randomized Trial. *Clinical Orthopaedics and Related Research*, 1648-1656.

Watanabe, S., & Yamaguchi, R. (1995). Intercultural Perceptions at the Workplace: The Case of the British Subsidiaries of Japanese Firms. *Human Relations*.

Wen, C., & Fang, J. (2012). Predicting potential respondents' decision to participate in web surveys. *International Journal of Services Technology and Management*, 16-32.

Witynski, M. (2011). *Behavioral economics, explained*. Retrieved from uChicago News: https://news.uchicago.edu/explainer/what-is-behavioral-economics

Wolf, C., Joye, D., Smith, T. E., Smith, T. W., & Fu, Y.-c. (2016). *The SAGE Handbook of Survey Methodology.* SAGE.

Wood, F., Siegel, E., & Feldman, S. (2008). Web Evaluation at the US National Institutes of Health: Use of the American Customer Satisfaction Index Online Customer Survey. *J Med Internet Research*.

Appendix A – Pattern Types for Collecting Data

Pattern and Purpose	Owner	Data Collection Frequency and Scope	Collection and Analysis Methodology	Discussion and Action Forums
Past Patterns: Captures a recent experience. > Intended to improve transactional experiences > Tracks experience goals and trends > Assesses impact of new initiatives > Identifies emerging issues Examples: Post-installation or customer service follow-up, new-product-purchase follow-up	Central group or functions	**Persistent:** > Electronic surveys linked to high-volume transactions or an ongoing feed-back system > Automatically trig-gered by the comple-tion of a transaction > Focused, short-cycle, timed data collection > Feedback volun-teered by users in online forums	> Web-based, in-person, or phone surveys > User forums and blogs	> Analyzed within functions, central survey groups, or both > Cross-functional issues directed to general managers > Strategic analysis and actions directed by general managers
Present Patterns: Tracks current relationships and experience issues with an eye toward identifying future opportunities. > Keeps a consistent yet deeper watch on state of relationship and other factors > Looks forward as well as backward > Used with more critical populations and issues Examples: Biannual account reviews, "follow them home" user studies	Central group, business units, or functions	**Periodic:** > Quarterly account reviews > Relationship studies > User experience studies > User-group polling	> Web-based surveys preceded by preparation in person > Direct contact in person or by phone > Moderated user forums > Focus groups and other regularly scheduled formats	> Initial analysis by sponsoring group > Broader trends and issues forwarded to general managers' strategic and operat-ing forums > Deeper analysis of emerging issues at the corporate, busi-ness unit, or local level
Potential Patterns: Targets inquiries to unveil and test future opportunities. Examples: Ethnographic design studies, special-purpose market studies, focus groups	General management or functions	**Pulsed:** > One-off, special-purpose driven > Interim readings of trends	> Driven by specific customers or unique problems > Very focused > Incorporates existing knowledge of cus-tomer relationship	> Centered within sponsoring group, with coordination by and support from central group

Appendix B – MINDSPACE Principles Definitions

PDF

MINDSPACE_Princip
le_Definitions_Kyra_

Appendix C – MINDSPACE Principle Selection

MINDSPACE_Princip
le_Selection_Kyra_Kc

Appendix D – Validity Assessment Framework

Validity Evidence	Questions	Answer
Researcher–participant relationship	1. Does the researcher or his/her sponsoring institution have a relationship with the respondent (e.g. no prior relationship whatsoever; alumni of the researcher's university)?	Health Corp and the respondent have a business-related relationship.
	2. What is the quality of that relationship (low vs. high level of trust; low vs. high level of commitment)?	This likely is a high level of trust and commitment.
	3. Is the pool of potential respondents drawn from a random sample?	Yes.
	4. Are potential respondents representative of the population to which the research seeks to generalize?	Not entirely, but the researcher is aware.
	5. Are the participants students taught by the researcher?	No.
Participant qualifications	1. Do potential respondents have first-hand knowledge of or experience with the phenomenon of interest?	Yes.
	2. In what ways are potential respondents representative of incumbents in this domain?	Yes.
	3. Are potential respondents qualified (possess requisite knowledge, skills and abilities)?	Yes.
Participant motivation	1. Is the participant being paid to complete the survey?	No.
	2. How much is the payment (compared with the country's minimum or average wage)?	It differs across the countries, but it is average.

	3. How many other surveys has the person completed in the prior seven days or month?	Unknown.
	4. Is there evidence of careful attention to detail (e.g. extensive attention checks)?	No.
Survey length and complexity	1. How cognitively demanding are the questions being asked on the survey (e.g. simple facts vs. complex judgments)?	They are very simple.
	2. How many items are on the survey?	1.
	3. How long does the average respondent take to answer all questions?	Approximately 3 minutes.
Number of times the survey is administered	1. Is the survey administered repeatedly?	No, unless it is not answered.
	2. If so, what is the expected relationship between frequent response and quality?	N/A.
Cultural and national context	1. What are the normative expectations for responding to surveys in the culture (e.g. organizational culture, national culture)?	It is very low (Vangeest & Johnson, 2011).
	2. Is the response rate in line with the normative expectations?	Averaging 29.5% (Vangeest & Johnson, 2011).
	3. Is there a reason to expect a coercive or forced response?	Yes.

Appendix E – Identified Points of Conversion

First login after psw reset

Customer effort

After at least 3 prescriptions created

Customer effort

2.1

2.2

2.3

2.4

POC Header 2.1

Based on your more recent experience. How helpful or unhelpful do you find the password reset process?
1. Very helpful
2. Helpful
3. Neither helpful nor unhelpful
4. Unhelpful
5. Very unhelpful

POC Header 2.2

How easy was the prescription process?
1
2
3
4
5

POC Header 2.3

How easy was the patient registration process?
1
2
3
4
5

POC Header 2.4

To what extent do you agree with the following statement: OrderManager's first order process was simple, straightforward, and painless.
1 - Strongly disagree
2 - Somewhat disagree
3 - Neither agree nor disagree
4 - Somewhat agree
5 - Strongly agree

POC Footer Basis

What's the main reason for your score?

After trial period end

2.7 7

2.8

POC Header 2.7

How helpful or unhelpful do you find the trial mode?
1. Very helpful
2. Helpful
3. Neither helpful nor unhelpful
4. Unhelpful
5. Very unhelpful

POC Header 2.8

How easy is the routine order process?
1
2
3
4
5

Feedback when survey is submitted

Thank you for taking part in our survey!

Appendix F – Logic Created by Health Corp

Survey_Logic_Kyra_K
oper.xlsx

Appendix G – Brainstorming Sessions BE

Unit non-response	Item non-response	Survey design	Unused biases

Unit non-response

Authority bias
- Logo of Fresenius apparent in survey

Endowed Progress Effect

Show a progress bar or amount of steps

Spark Effect

"This will only take 5 minutes"

Analysis Paralysis

Avoid same feedback all the time

Vary message and "reward"

Peak-end rule

Social facilitation / proof

In the end show x amount + 1 more (you!)

(show how many others have already taken the survey)

Goal Priming

Show the value for their patients if use

Feedback Loops

Curiosity Effect

Do you want to improve the experience of hundreds of patients?

Goal Priming

Curiosity Gap

Timing

Ownership bias + Reciprocity

Ownership Bias

Hey _____! Give us feedback so we can improve your experience

Item non-response

Progressive disclosure

Limited Choice

Analysis Paralysis

Zeigarnik Effect

If user quit the survey, we save it for later

2 questions to finish! will only take 1 min

NurseApp: Survey task in the list

Smaller surveys, easy to fill in

Your opinion is important for use (make them feel special)

Anonymous survey

Survey design

Unused biases

Foot in the door

Collection bias

Spacing effect

Default effect

Storyteller Bias

Commitment

Appendix H – Survey Questions Mock-Ups

Appendix I – Brainstorming Session Survey Design

Appendix J – Survey Starting Screens
Non-BE version

OrderManager

We welcome your feedback

Help us improve your experience
by taking our short survey.

Give feedback

Not this time

Authority Bias

Company Logo

Company Name
welcome your feedback

Help us improve your experience
by taking our short survey.

Give feedback

Not this time

Social Influence

Already 98 other nurses have given us feedback about their experience

WIl you give us your feedback, too?

Give feedback

Not this time

Loss Aversion

OrderManager

Please take part in our short survey

By not taking part of the survey, you will lose your opportunity
today to improve your usage of the app.

Give feedback

No, I'll miss out

Appendix K – Survey Questions
Non-BE version

Authority Bias

Company Logo

How easy is the prescription process?

○ Very Difficult

○ Difficult

○ Neither easy nor difficult

○ Easy

○ Very Easy

Any recommendations to improve the process? - Optional

[ ]

Submit

Skip feedback

Social Influence

How easy is the order process?

○ Very Difficult

○ Difficult

○ Neither easy nor difficult

○ Easy

○ Very Easy

Any recommendations to improve the process? - Optional

[   ]

Submit

Skip feedback
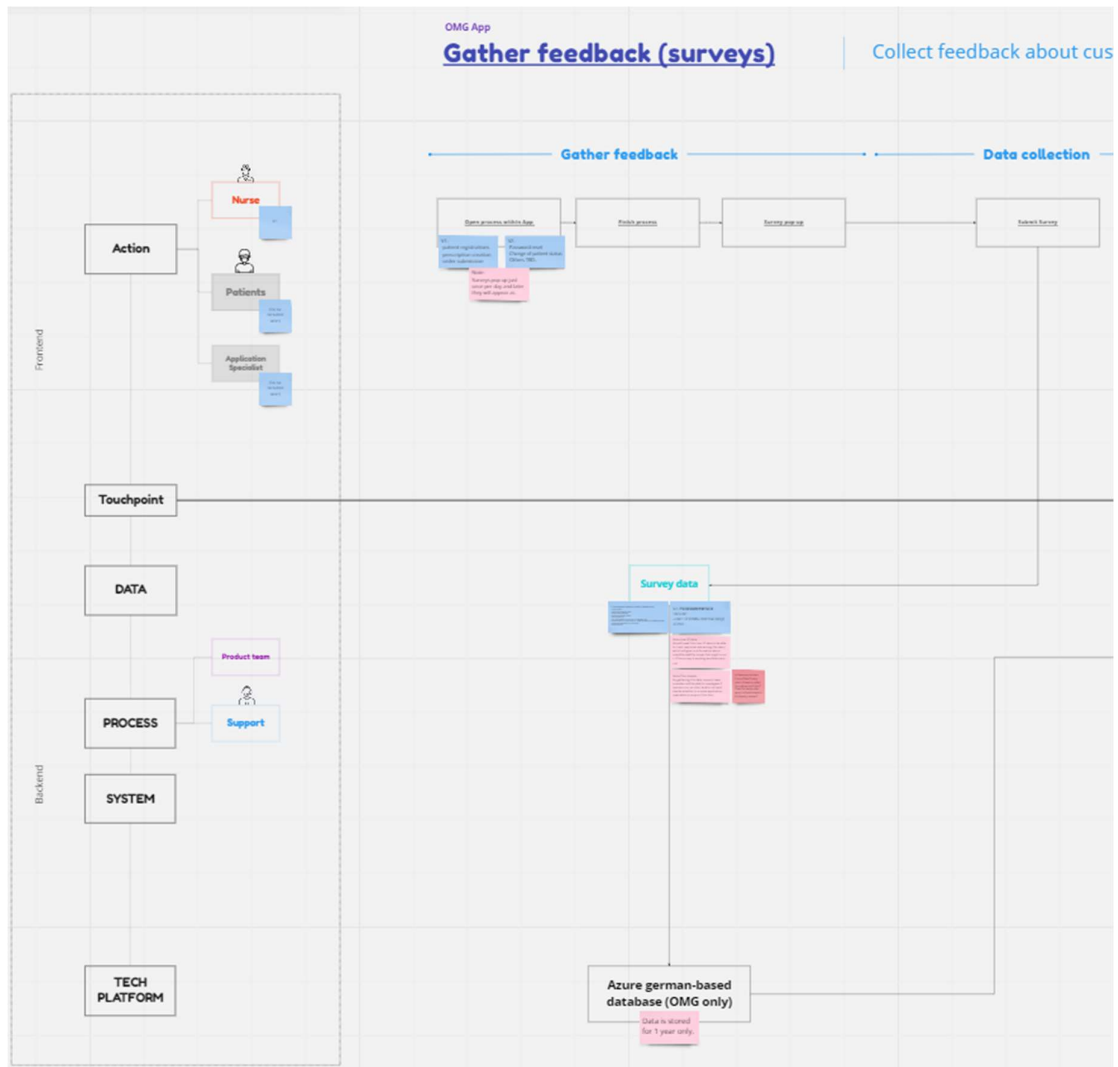
Loss Aversion

Appendix L – Survey Thank You Screen

Appendix M – Excel Report Survey Data

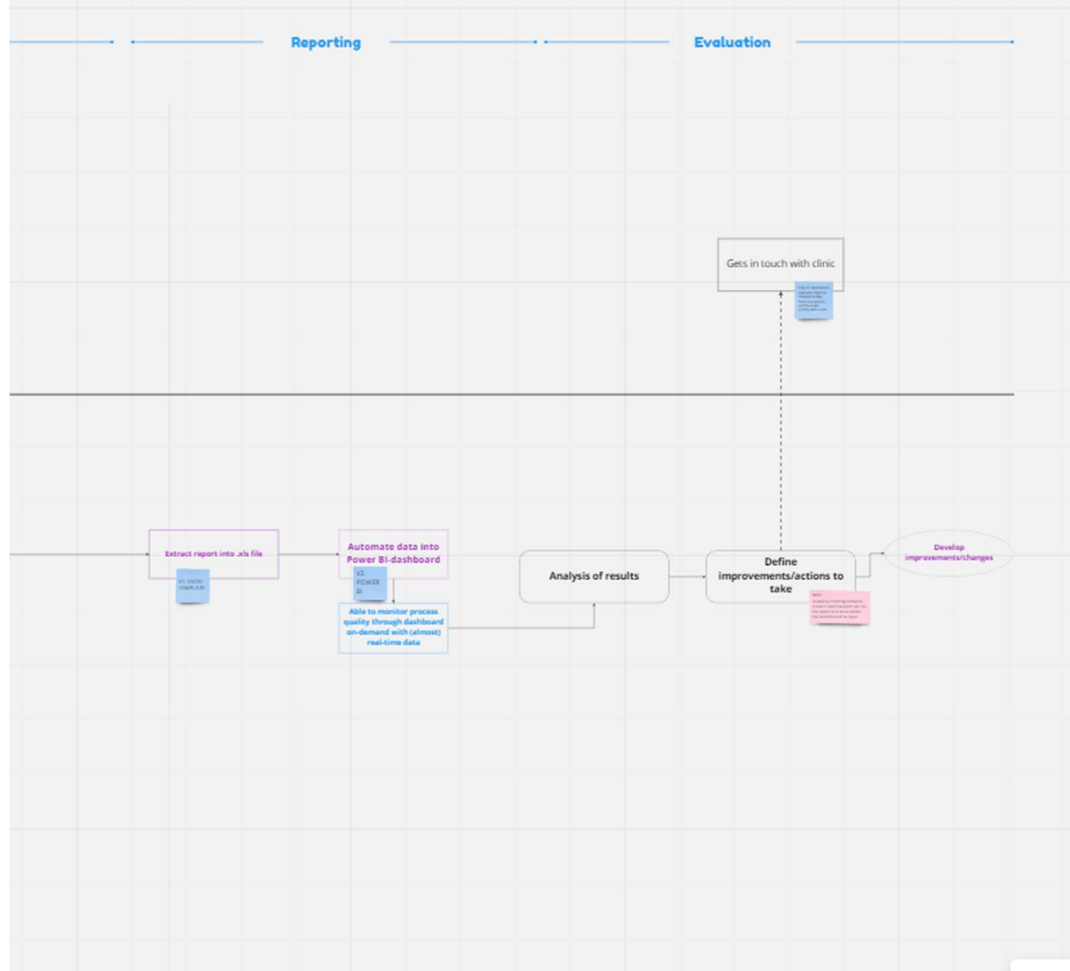
User_Feedback_Rep
ort_Kyra_Koper.xlsx

Appendix N – Survey Data Gathering Process

...stomer experience among nurses and patients through pop-up surveys.

Appendix O – Chi-Squared Table

Degrees of freedom (df)	Significance level (α)							
	.99	.975	.95	.9	.1	.05	.025	.01
1	-------	0.001	0.004	0.016	2.706	3.841	5.024	6.635
2	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210
3	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345
4	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277
5	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086
6	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812
7	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475
8	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090
9	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666
10	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209
11	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725
12	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217
13	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688
14	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141
15	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578
16	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000
17	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409
18	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805
19	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191
20	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566
21	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932
22	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289
23	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638
24	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980
25	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314
26	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642
27	12.879	14.573	16.151	18.114	36.741	40.113	43.195	46.963
28	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278
29	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588
30	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892
40	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691
50	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154
60	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379
70	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425
80	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329
100	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116
1000	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807

Taken from (Scribbr, 2022).