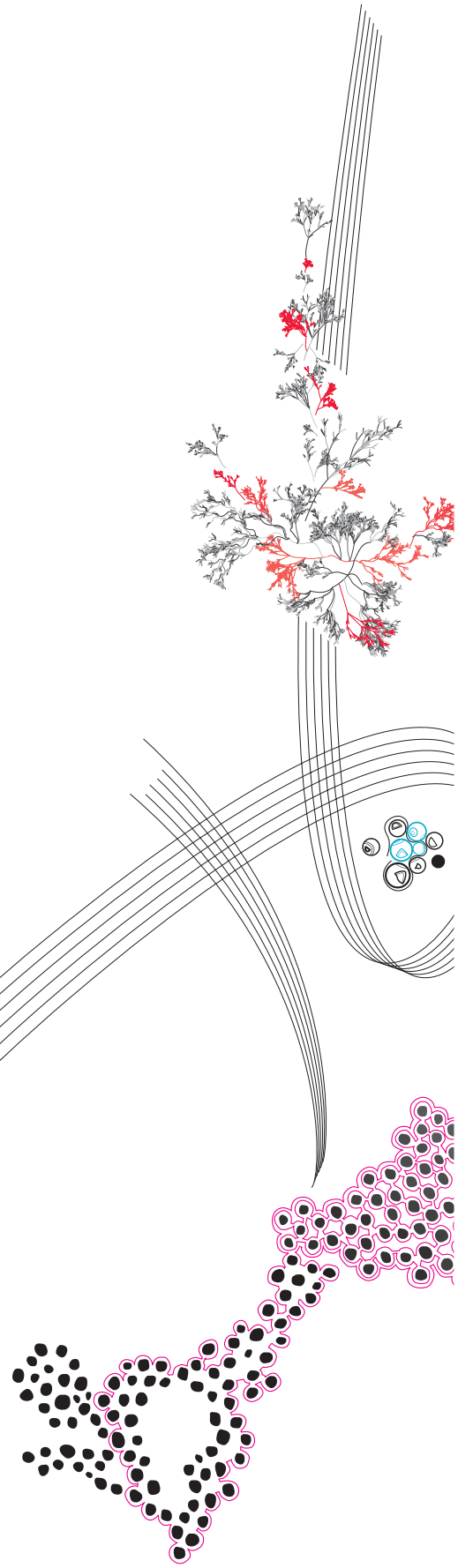

M. Sc. Interaction Technology

Master Thesis

Data-Driven Decision Support and Awareness of Risk Related Biases

Saskia Hustinx

Supervisors:
dr. R. Guizzardi,
dr. M. Theune

September, 2024

Faculty of Electrical Engineering,
Mathematics and Computer Science,
University of Twente

UNIVERSITY OF TWENTE.

Abstract

This thesis examines the use of a dashboard in combination with a decision support screen in order to debias financial project management decisions. The research integrates findings from choice architecture, dashboard design, and debiasing into a combined system, in order to create bias awareness and steer the user towards reflection. The developed system was first tested using preliminary user tests, ensuring the usability of the components. The updated system was then validated through qualitative user tests, semi-structured interviews, and a questionnaire on user acceptance comparing two alternative system designs. The questionnaire results show a slightly higher rating of system benevolence for a transparent system design compared to a covert one, but no significant differences in other dimensions, such as user trust or perceived manipulation. The research results of the user tests indicate that while the use of the system seemed to enhance bias awareness, the applied debiasing strategies did not necessarily lead to a reflection or reconsideration of biased decisions.

Keywords: decision support, debiasing, nudging, bias awareness, project management

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

The financial decision making process is a highly complex process that is influenced by an array of different factors and circumstances. Historically, researchers based their decision models on mathematical theorems, which assumed that human decision makers act as purely rational agents (*homo economicus*), that always select the most profitable or mathematically optimal option [66, 43].

This was later revised with the rise of behavioural economics and concepts such as *prospect theory*, which was pioneered by the research of Kahneman and Tversky in the 1970s [35]. This new wave of research proved that in reality, most financial decisions are not purely rational or optimal, but highly subjective and intuitive. Instead of assuming perfect rationality, the field of behavioural economics also takes into account the impact of human factors like emotions or personality on the decision making process [64]. When confronted with complex scenarios combined with uncertainty, people tend to simplify decisions by using heuristics (mental shortcuts), which then in turn lead to *cognitive biases* [35, 36, 64]. These biases occur in a wide range of tasks, and they can range from data processing biases that occur while consuming new information, to personality- or circumstance-based biases like overconfidence [39].

Due to the correlation of cognitive biases with uncertainty and risk, they are very prevalent in the financial sector, for example in project-management, investment, and entrepreneurship [1, 11]. Although these biases and heuristics do not always lead to negative outcomes, it is important to keep their influence in mind when developing decision support systems and decision making interfaces. Within the business domain, research

indicates that cognitive biases can lead to escalated project investments and irrational investments, which shows the need to monitor and anticipate such influences on the decision making process [38].

To counter the impact of cognitive biases on decision outcomes, a number of debiasing strategies have been developed and applied with varying degrees of success [40]. An example of such a strategy is bias awareness, in which users are made more aware of possible biases before finalizing decisions. Research shows that creating awareness about the existence and workings of biases and heuristics can already improve the decision quality and aid the decision maker in their process [28, 55]. Besides such educational strategies, another approach to debiasing is the use of nudges. This describes the use of certain design principles and interventions to steer the user towards a certain outcome or process, which can be used to debias decisions [57, 10].

Another vital part of debiasing is providing the user with well-presented, high-quality data to improve the decision quality and ease the cognitive strain when faced with complex scenarios. Without well-structured and sorted data, decision makers are more likely to fall victim to the previously mentioned biases, as well as certain cognitive limitations, such as decision fatigue and information-overload [57, 40].

One way to present large amounts of data in a structured and concise manner are *dashboards*. Within the business domain, dashboards have become increasingly popular as a tool for displaying and analyzing data over the past years, particularly due to the ever growing amount of collected data. Dashboards are meant to help structure and condense large amounts of data, making the data more accessible for the end-user. Especially in fast changing and complex environments, they are frequently used to monitor incoming data and performance indicators, which makes them a common tool in project management and finance [69].

While existing literature extensively covers guidelines and design principles for such strategic and organizational dashboards, as well as strategies for debiasing and bias awareness, there is very little on the use of dashboards for debiasing. This shows a need for further research on the combination of dashboards and debiasing strategies, as dashboards are already a popular tool for business and project management, and could therefore be easily integrated into existing workflows. This thesis proposes a tool combining the previously mentioned technologies into a strategic bias awareness dashboard.

1.2 Research Questions

This thesis proposes the combination of findings from the fields of dashboard design, decision architecture, and debiasing strategies into an integrated bias awareness and decision support tool. The goal is to support project managers in their decision making process, by creating awareness about potentially biased decisions and their outcome, as well as giving them a clear and structured overview of necessary data when making new decisions.

This thesis is part of a larger project aimed at providing decision making support considering cognitive biases. As part of the research, an ontology defining cognitive bias has been developed [13]. This ontology serves as basis for the development of a machine learning algorithm to detect cognitive biases [54] and for the dashboard developed in this thesis.

To limit the scope of the project, the focus lies on risk related biases (risk seeking and risk avoiding behaviour) that occur within go and kill decisions in project management. This describes scenarios, in which a project manager is asked to decide on whether to continue or stop a project when it is being re-evaluated.

Based on these goals, the main research question that will be addressed in this thesis is ***how to raise bias awareness for risk related biases in project management contexts?*** This is split into two sub questions addressing individual parts of the proposed tool. Sub question (a) will go into how to combine strategic dashboard design with debiasing and bias awareness. Sub question (b) will look into how decision architecture and the nudge transparency impact the decision process and user acceptance in this specific context of project management.

RQ: *How to raise bias awareness for risk related biases in project management contexts?*

> **1)** *How to visualize past decisions and current project data in a project management context to help create bias awareness?*
>
> **2)** *How does the transparency of nudging in decision support tools affect the decision making process and user acceptance in project management contexts?*

1.3 Methodology

The goal of this research is to integrate elements from debiasing, choice architecture, and dashboard design into a combined solution. First, a brief requirement analysis is conducted. Based on the gathered requirements and a review of related literature, a first design is then proposed. To validate the functionality of this design, a series of usability tests are done using a clickable mock-up. These usability tests follow a 3-part structure: a free exploration of the prototype, a series of tasks to complete with the prototype, and a final interview. The interview questions are loosely based on the PSSUQ (Post-Study System Usability Questionnaire) [42] and System Usability Scale (SUS) [8] questionnaires, but adapted to gather more qualitative results. In addition to the interview results, a number of usability metrics, such as the completion rate and the time spent on task are collected. Based on the gathered feedback and results, an improved version of the design is created. This design is then implemented using the specified technologies determined during the requirements phase of the project. This finalized design is then validated through qualitative user tests, semi structured interviews, and a questionnaire on user acceptance, to gain an insight into the effectiveness and usefulness of the proposed system.

1.4 Thesis Outline

First, an overview of the relevant literature for each of the related subject areas will be given in Chapter 2. The selected methodology and technologies, as well as the created artefacts for each stage of the process (such as the designs and prototypes) will be presented in Chapters 3 and 4. The preliminary usability tests will be presented in Chapter 5. The validation methods and gathered results will be presented and discussed in Chapters 6. Finally, Chapter 7 will discuss the gathered results, and Chapter 8 will give a conclusion to this research and the addressed research questions.

Chapter 2

Background and Related Work

This chapter will summarize the relevant literature on the different components needed for visualizing and debiasing project management decisions.

The first section will give a brief overview on behavioural economics and the underlying processes that lead to biases in financial decision making. The second and third section will go over available debiasing strategies and approaches, as well as guidelines for creating decision architectures. The last section will summarize the most important literature on strategic dashboard design and visualizations.

2.1 Behavioural Economics and Cognitive Biases

Normative vs. Descriptive Economic Theories

Behavioural economics are a field of research that gained popularity in the 1970s and 80s, as an alternative to *normative economic theories*. These normative approaches that were prevalent until then, depict a rational, mathematically optimal decision making process, in which the decision makers make the best possible decisions [66, 43]. Most of these normative theories, like the *expected utility theory*, are applied to optimize decision problems, therefore choosing the option with the highest expected outcome or utility [63].

The approach of behavioural economics on the other hand is a descriptive one, meaning that it aims at explaining how and why decisions are made in practice. It includes elements from psychology and cognitive sciences to explain how factors like emotion and personality impact the human decision making process [64, 36, 56].

One of the most prominent theories of behavioural economics is the *prospect theory* [35]. The theory highlights three main points that contradict the expected utility theory: The

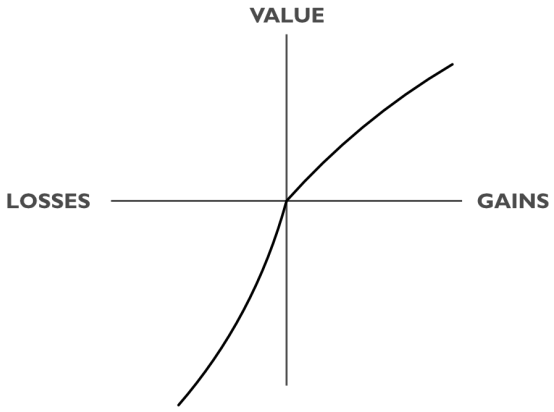

FIGURE 1: Hypothetical Value Function - Kahneman and Tversky

perception of losses and gains, the framing (wording) of messages, and the perception of certainty. It shows, for example, that gains and losses are not perceived as equal: This means that a loss of a certain amount is perceived as more painful that a gain of the same amount would bring joy, as shown in the utility function by Kahneman and Tversky in Figure 1 [35]. The graph shows that the perceived value function is steeper for losses than it is for gains.

An example of this, in combination with the importance of the framing of options can be seen in the following experiment conducted by Tversky and Kahneman [65]:

> **Problem statement:** *Imagine that the U.S. is preparing for the outbreak of an unusual disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Which of the two programs would you favor?*
>
> **Scenario 1:**
>
> - If Program A is adopted, 200 people will be saved. [72 percent chose this option]
> - If Program B is adopted, there is 1/3 probability that 600 people will be saved, and 2/3 probability that no people will be saved. [28 percent chose this option]
>
> **Scenario 2:**
>
> - If Program C is adopted 400 people will die. [22 percent chose this option]
> - If Program D is adopted there is 1/3 probability that nobody will die, and 2/3 probability that 600 people will die. [78 percent chose this option]

Although the odds and outcomes of the two scenarios are exactly the same numerically, the participants largely prefer the first option in scenario one and the second option in scenario two. This higher level of risk-taking in the second scenario can be attributed to the framing of the options (the wording of 'amount of people saved' vs. 'amount of people died') and the more negative attitude towards losses, as illustrated in the value function in Figure 1.

These examples show a tendency for people to act more risk seeking when presented with a negative frame (losses) in contrast to being more risk averse when presented with a positive framing (gains). Framing, among other biasing factors (see next subsection), can therefore shift the decision makers perception of risks and unconsciously lead to more risk seeking or averse behaviour [67].

Heuristics and Cognitive Biases

Research in behavioral economics demonstrates that the decision making process in real-life scenarios is rarely optimal or fully rational: Decisions often need to be made with high degrees of uncertainty, sub-optimal data, or with a general lack of information. In such cases, people tend to apply *heuristics* (mental simplifications or shortcuts), when making a decision [20].

For the most part, the application of heuristics is an unconscious process that produces passable results in stressful or uncertain situations. This is useful from an evolutionary perspective, as it preserves energy and leads to quick results under stress or in dangerous situations [20]. In most cases, the results of such intuitive decisions are good enough, but in some cases the heuristics can reduce the decision quality and lead to certain errors known as *cognitive biases* [64].

Cognitive biases are broadly defined as systematic human errors and inconsistencies that occur in the perception and evaluation of decision problems [65]. Nowadays, there is research on over two hundred known cognitive biases, that can range from data-processing errors to more general, personality-based biases like overconfidence [20, 24]

An example of such a bias is the *sunk cost fallacy*: It describes a phenomenon of escalated commitment, meaning a tendency to justify further investments in a failing project after having previously invested in it [24].

The specific biases that occur in decision making can vary per decision domain. Some of the most common cognitive biases in project management include the escalation of commitment (sunk-cost fallacy), overconfidence, and optimism bias [24].

Dual Process Theory

Research has shown that cognitive biases and systematic errors do not appear randomly, but that they occur the most frequently when making quick and intuitive decisions [36, 64]. That sets this kind of intuitive decision making apart from situations in which decision makers reason more deliberately before making a decision [33]. To further distinguish between these two different modes of decision making, behavioural economics use a *dual-process theory* (System 1 and System 2 thinking) to differentiate between intuitive and deliberate decision making [33].

System 1 (Intuition) is used to describe a quick and impulsive way of decision making, which is often more prone to systematic biases and errors [33]. Cognitive psychologists estimate that humans operate roughly 95% of the time in the more intuitive mode of System 1 processes [14].

System 2 (Reasoning) is used to describe a more slow, deliberate, and carefully reasoned way of making decisions [33]. Due to the more conscious nature of System 2 (in comparison to the more reflex-like nature of System 1), it is more likely that judgement errors are caught and corrected while making decisions [14].

Decision makers are generally able to switch between the two modes of thinking when the need arises, but tend to automatically default to the less costly System 1 whenever possible [14]. This switch between System 1 and 2 thinking is the key to almost all debiasing strategies: Most approaches for debiasing attempt to bridge the gap between the two systems by trying to steer the user towards reflection and deliberation [34, 57, 40]. Some concrete approaches for debiasing and bias awareness are summarized in the next section.

2.2 Debiasing Strategies

Past research has shown that cognitive biases do not appear randomly, but that they are somewhat predictable and usually follow certain patterns of behaviour [36, 64]. Therefore, a significant part of the research on cognitive biases nowadays focuses on the development of strategies to counteract them. This section discusses some of the most prominent strategies and approaches to debiasing found in literature.

In general, the key to most debiasing strategies is overcoming the gap between the previously mentioned System 1 and System 2 thinking. This is done by guiding the decision maker towards a more careful reflection of the decision, and away from the intuitive use

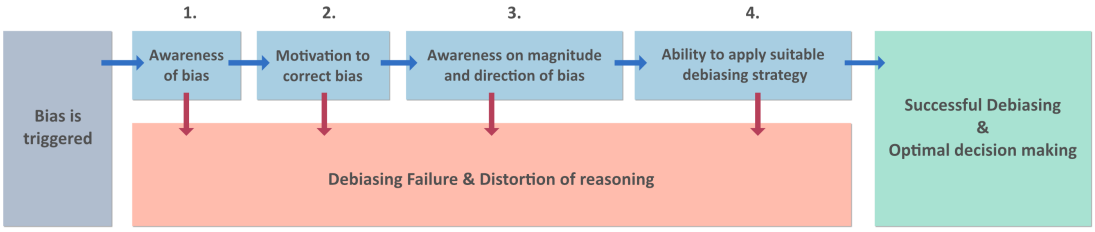

FIGURE 2: Debiasing Process, adapted from [14] and [68]

of heuristics that usually lead to biased decisions [34, 57, 40]. According to Kahneman and Frederick, in order to correct a flawed intuition, them decision maker needs to reason carefully and reflect at least briefly before finalizing their decision [34]. Since the brain tends to automatically and unconsciously fall back on System 1 thinking, the change towards reflection and reasoning (System 2) needs a conscious effort in form of debiasing interventions [14].

Croskerry et al. [14] and Wilson and Brekke [68] describe successful debiasing as a series of four consecutive steps, as seen in Figure 2: (1) awareness of bias, (2) motivation to correct bias, (3) awareness on magnitude and direction of bias, (4) ability to apply suitable debiasing strategy.

The first step towards successful debiasing is being aware of the biases, in combination with a motivation to correct them. This is followed by having an awareness of the severity and impact of the biases, so that an appropriate debiasing strategy can be applied. The completion of all four steps leads to successful debiasing and optimal decision making. A failure at any stage of this process, however, can result in failed debiasing and distorted reasoning. Fischoff [23] describes a similar set of strategies, including offering warnings about possible biases, describing the bias and its influence in detail, providing the user with feedback, and offering appropriate training interventions.

Classic literature on the topic of debiasing usually divides debiasing strategies into two categories: *Internal and external* [40, 23]. These categories are occasionally also referred to as '*modifying the user*' and '*modifying the environment*' [57].

Internal strategies (modifying the user) mainly focus on providing the decision maker with suitable cognitive strategies to avoid them falling back on the biased heuristics. The main focus of *external strategies* (modifying the environment) lies on optimizing and improving the decision environment to avoid triggering the biases in the first place, or aiding the decision maker through the use of tools and technology [40, 57]. Both of these

categories share the same basis, that debiasing usually needs an intervention, since the decision makers are most likely not able to debias themselves without external input [40].

Besides the mentioned internal and external debiasing strategies, there is also a more lightweight alternative called *organizational cognitive repair* [57]. Such a cognitive repair (sometimes called a motivational repair) is a small habit or ritual, such as a often repeated proverb, that aims at improving the decision making process at a low cost. One such example of a cognitive repair is the implementation of short checklists ('*before you submit x, check y and z*') [57].

A more direct way of steering the decision maker towards a certain direction or outcome is called *nudging*. A nudge is a modification of the decision environment that is meant to influence the decision making process in a desired way. Such nudges are usually counted as external debiasing strategies (modify environment), and can potentially be very successful for debiasing [57]. Nudging, as a special form of debiasing, will be discussed in more detail in Section 2.3.

In the next two subsections, the most important debiasing approaches for each of the two main categories (internal and external) will be discussed.

Internal Strategies (Modifying the decision maker)

The key point of internal debiasing strategies is the education and training of the decision maker. Larrick [40] differentiates between three different types of user training: Training in rules, training in biases, and training in representations.

Training in rules aims at equipping the decision maker with normative strategies specific to the use case, or with more general cognitive tricks that can be applied to mitigate biases [40]. Such training could include a detailed strategy on how to approach a certain type of problem, or promote the use of specific cognitive strategies that try to evoke reflection. An example of such a cognitive strategy is 'consider the opposite' or 'imagine it fails'. Using this strategy, the user is encouraged to consider a failure of the current plan or preference and to list the pros and cons of that alternative outcome [57].

Training in biases aims at educating the user about cognitive biases in general, how they work, and how to recognize them [40]. This is especially important, since the users need to be able to recognize situations in which debiasing interventions are needed. This education creates the required foundation of bias awareness, understanding, and sometimes motivation, that is needed for successful debiasing (as seen in Figure 2).

Training in representations is meant to improve the statistical reasoning of decision makers [40]. Kahneman and Tversky have proven that a lot of cognitive errors and biases stem from the fact that humans are intuitively not very good at statistics, especially when numbers are presented as probabilities [36, 34]. According to their research, decision makers can make better use of relative frequencies (e.g. 1 in 5) than they can with probabilities (0.2) or percentages (20%) [34]. This can either be used externally, by converting probabilities to frequencies in visualizations (see next subsection), or internally, though training. In such training, decision makers could be advised to mentally convert and compare between the different notations of values when considering the options of a decision.

External Strategies (Modifying the environment)

The main point of external strategies is to adapt the decision environment in a way that improves decision making. Larrick [40] and Soll et al. [57] discuss various debiasing strategies that can be roughly sorted into five categories: Incentives, social factors, framing, technology, and nudging (see next section).

The use of *incentives* is based on the assumption that the more analytical thinking of System 2 will eventually kick in, if the stakes (like through a monetary incentive) are high enough. According to Arkes [4], having higher stakes leads to higher efforts and a more careful consideration of the underlying decision data. This, however, only works for some biases, for example ones that are based on the neglect of data or insufficient attention [57]. For other biases that occur because of faulty heuristics and strategies (like the framing effect or overconfidence), traditional incentives do not work well, as it just leads to the same biased behaviour with higher levels of enthusiasm ("think harder but not smarter" [40]) [4].

A slightly different form of incentive is *accountability*, where decision makers are either publicly held accountable for their choices, or where they are under the impression that they will have to justify their decisions in the future. This acts as a special form of incentive, by introducing a social factor into the decision making process. Similar to monetary incentives, accountability leads to greater effort and more thorough information usage, by increasing the cost of failure [57]. (In this case a social cost instead of a monetary one.) Accountability as a reflection-inducing strategy has proven successful for a handful of biases, including the sunk-cost fallacy. Similar to monetary incentives however, it is largely unsuccessful for biases, where the main problem is not a lack of effort or attention [57].

Another group of external debiasing strategies include *social factors*. Similar to accountability, these strategies are based on a range of group- and social-based behaviours. One such behaviour that can be used accordingly, is herding [57]. This describes the human desire to conform to social norms, and to adhere to what is perceived to be the group standard. This can be used in decision making, by conveying which option other people have chosen, what the average choice is, or by applying group decision making. Decision making in groups can create synergies and more well-reasoned decisions, as well as compensate for strong outliers by averaging the choices. In some cases however, social factors can also have detrimental effects, for example when people hide their true opinions while in a group-setting [57].

Framing, as mentioned earlier in Section 2.1, is the specific way in which messages are worded and values are conveyed to the user. This can be used in debiasing, by either framing messages as neutral and accessible as possible, or by framing the messages in a more leading way [4, 34]. The latter option counts as nudging, and will be discussed in more detail in the next section. An important factor of framing is the way that gains and losses are presented, as that tends to evoke strong feelings within the user [65]. For example, concatenating gains (he won 10€, and then he won 20€) generates more joy for the user than seeing the total outcome (he won 30€ in total) [4]. Another example that uses re-framing to improve the decision environment is presenting values as frequencies (e.g. 1 in 5) or visual representations (diagrams etc.), rather than probabilities (.20) as they are easier to process [34].

Finally, the last category of external debiasing strategies is the *use of technology*. This includes the use of decision models, decision support systems, and visualizations [57]. In more repetitive or simple scenarios, the use of mathematical decision models can be used as guidance and advice, by generating the mathematically optimal outcome. This approach however, is not suitable for more complex scenarios with fast-changing variables [40]. Another element within this category is the use of visualizations and information screens to aid the decision making [57]. It is proven that visual representations influence the interpretation of data, and can therefore be used to aid debiasing [20]. This is especially useful, since visual information and numerical displays are generally easier to process than equivalent textual data [20, 40]. Another approach that typically combines the use of visualizations with decision models are decision support systems (DSS). These systems have a lot of potential for debiasing, because they can shape the way that decision data

is presented to the user [40]. This kind of design is also known as choice- or decision architecture, and will be discussed in Section 2.3.

Selecting a Strategy

When it comes to selecting appropriate debiasing strategies, there is a number of factors that should be considered, such as the environment, decision complexity, and the target group.

First, there is a *temporal component* to decision making and debiasing, that should be taken into account when choosing a strategy [15]. According to Bhandari et al. [6], a successful debiasing approach for financial decision making should cover all three temporal dimensions (past, present, future). They propose a system that examines past decisions and their outcome (past), aids the current decision making process by providing structured and relevant information (present), and warns the user about possible biases that might occur and how they might impact the overall situation (future).

Besides the temporal aspects, the effectiveness of different debiasing strategies is always dependent on the specific *use case and the target group* that is supposed to use the system or intervention. Especially in cases where a certain intervention is implemented top-down from the executive level to the employees, it has a higher tendency to fail [57].

Another factor is the *policy maker* (who selects and implements the interventions) themselves. Soll et al. [57] recommend that the more questionable the competence level or benevolence of the person designing or implementing the interventions is, the more appropriate internal strategies (e.g. training the decision maker) become.

The specific *qualities of the decision* itself should also be considered when choosing a strategy: For more complex decision scenarios that occur somewhat infrequently and are subjective to the decision maker, an internal approach (modifying the decision maker and providing them with normative strategies) might be most effective [57]. On the other hand, if a decision occurs more frequently, is objective, and less complex in nature, external approaches like a decision model might be more appropriate [57].

The last factor that should be taken into account is *decision readiness*. This describes a state, in which System 2 is capable to function at full capacity and without issues [57]. According to Soll et al. [57] there are three factors that have an influence on this: Fatigue or distraction, emotional (visceral) influences, and individual differences (such as training,

personality or intelligence). The selected debiasing strategies should aim to facilitate this decision readiness for the user, if it is not already provided [57].

Regardless of the selected debiasing strategy, a major factor to consider is the *user acceptance* of such interventions. This is generally considered a difficult topic, as many people are reluctant to accept decision aid and reconsider their past decisions [40]. Past research has shown that even when people are made directly aware of their inconsistent or biased decisions, many are not willing to change their answers [4]. This shows how difficult debiasing can be, especially when there is little intrinsic motivation [4, 14]. In most cases, successful debiasing only works if the user is motivated to improve and debias their decision making [14, 40]. Especially in corporate settings, a distinction needs to be made between a rather unmotivated compliance to a certain strategy, or the internalization of beliefs [57, 40]. In such settings, a debiasing strategy is the most effective if it can be integrated into the existing process, and if it is internalized and seen as a valuable addition by the employees themselves.

2.3 Decision Architecture and Nudging

When making decisions, the specific way in which the available choices are presented to the decision maker can have a significant impact on the outcome (e.g saying something has a 50% change of failure rather than a 50% chance of success). The design of such a decision environment is called *choice or decision architecture*, and the corresponding designer is called a *decision architect* [57, 60].

When a decision architecture is engineered in a way that supports a certain outcome, that is generally called *nudging* [57, 61, 10]. The term *nudge* was originally introduced by Thaler and Sunstein [60] as an expression for interventions that use the decision architecture as a way to steer people's behavior in a predictable way without limiting their choices or offering monetary incentives. According to them, nudges should only rely on psychological principles to steer the decision making, and should be easily avoidable by user, if they desire so [59, 60]. A classic example of a nudge is the use of defaults: If a company sets up their retirement-plan in a way that enrols employees automatically if they do not opt-out, the enrolment rate is significantly higher than if they actively need to opt-in [57]. In such a case, the choices for the decision maker are not limited (they can opt out), but the decision environment is structured in a way that nudges the user towards a certain outcome.

Nudging is typically considered an external debiasing strategy if it modifies the environment to steer the user towards rationality [57], although there is sometimes no clear distinction as to where reflective debiasing stops and the more direct nudging begins. Figure 3 shows an adapted visual overview the mentioned debiasing interventions in relation to the decision phase they target, and their categorization according to Soll et al. [57] and McKenzie et al. [45].

The principle of nudging has faced some criticism over the years, as it can be seen as coercive or paternalistic towards the decision maker [10, 45]. The key factor in this discussion is the question of who the nudge primarily benefits: the decision maker, the society, or the nudger (decision architect) themselves [10]. To distinguish benevolent nudges (focused on the decision makers welfare) from marketing strategies (focused on decision architects profit), Thaler and Sunstein introduced the notion of *libertarian paternalism* [59]. They argue that some forms of paternalism should be acceptable when designing a choice-architecture, if they aim at improving the decision makers' welfare and rationality [59]. This is founded on the idea that in reality, no decision architecture can ever truly be

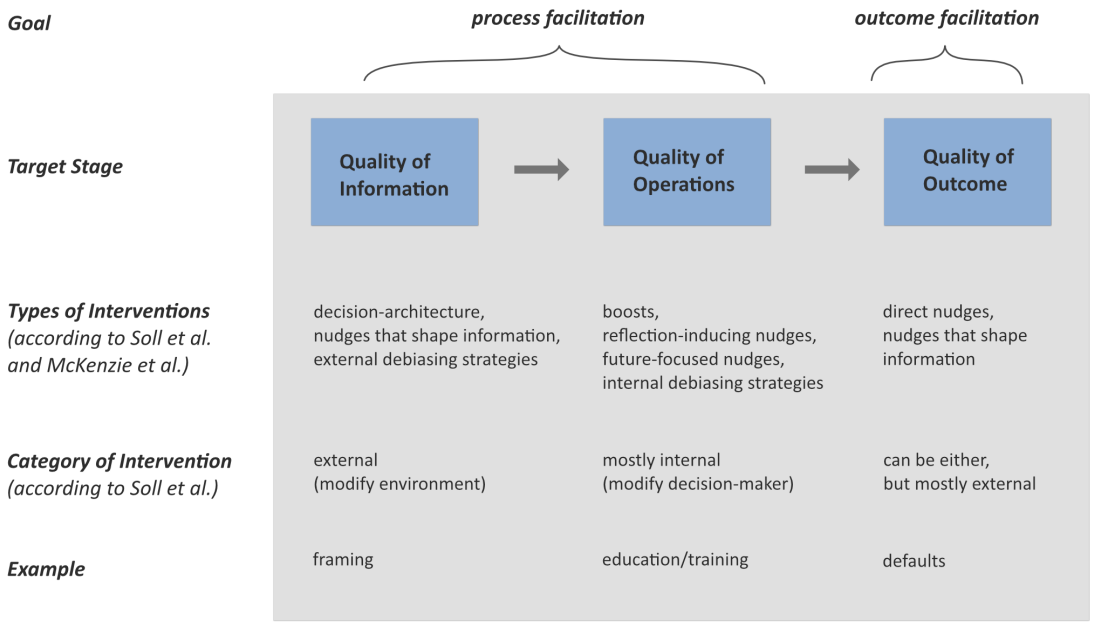

FIGURE 3: Overview of Debiasing Strategies per Target Stage, [57, 45]

neutral. It is argued that every design decision, from the choice of words to deciding against certain interventions (like defaults), has an influence on the decision making process, even if that is not intended [59, 60, 32]. According to the theories of libertarian paternalism, nudges only steer decision makers in directions that they would have chosen themselves, if they were rational and not impacted by biases [10, 59].

Attempts have been made to make nudging less paternalistic, for example by providing transparency to the user about the used cues, underlying intention, and the desired outcomes of interventions [45]. This will be briefly discussed at the end of this section.

Nudging Strategies

As for the actual nudging strategies, Soll et al. [57] categorize the most popular approaches into four main groups: (1) defaults, (2) nudges that induce reflection, (3) nudges that induce future-focused thinking, and (4) nudges that kindly shape information.

Providing *default options* is considered one of the most effective strategies for nudging. It makes use of the decision makers inherent inertia and tendency to preserve the status-quo, in which people tend to choose the option that takes the least effort (path of least resistance) [57, 61]. According to Soll et al. [57], algorithmically chosen *smart defaults* are an option that can be especially valuable for debiasing financial decisions, as it provides high-quality,

credible defaults. A factor that should be taken into consideration when working with defaults is that users tend to perceive pre-selected defaults as implicit recommendations [45]. This can backfire, especially if the quality of the defaults is sub-optimal, or if the decision makers question the motives of the policy-maker [62, 45]. Research has shown that people that have been presented with high-quality defaults in the past are more likely to follow defaults in the future, than people who were exposed to random, or sub-optimal defaults in the past [62].

Reflection inducing nudges try to shift the user to rely less on their instinctual response (System 1) and steer them towards reflection (System 2 thinking). Examples of this are planning prompts, planned interruptions, and active choice [57]. Planning prompts require the user to complete a certain decision or task at a pre-determined moment. This helps overcome inertia and forgetfulness [57]. Planned interruptions (e.g. making a required resource temporarily unavailable) are meant to slow down the decision making process and prompt deeper reflection while the task is on hold [57]. Finally, requiring decision makers to make an active choice (in comparison to mindlessly accepting a default or avoiding the choice altogether) is also meant to nudge the user towards a more careful reflection [57].

Nudges toward future focused thinking attempt to nudge the user by removing the spontaneous element of decision making that often leads to intuitive decisions (System 1). Examples of this are pre-commitment, choosing in advance, and temptation bundling [57]. Choosing in advance and committing to that choice are two strategies that are meant to help with the near-sightedness of spontaneous decision making: decisions that are made in advance improve the decision makers ability to think abstractly about a topic and increase the selection of long-term benefits (that might not look as attractive on a short-term basis) [57]. Temptation bundling is the combination of positive (instantly gratifying activities) and negative influences (long-term benefits, but short-term costs), in order to increase long-term commitment [57].

Nudges that kindly shape information are meant to promote better decision making by presenting the required data in a comprehensible and compelling format [57]. The most important strategies for shaping the information are framing, transforming scales (frequencies vs. probabilities), conveying social norms, categorizing, and using visualizations [57, 32]. This largely overlaps with the previously mentioned external debiasing strategies in Section 2.2, as the lines between active nudging and the intentional framing of information are not very distinct. Here, it depends on the policy-maker and their ultimate goal, to

decide on the specific framing of messages, and whether it is meant to purposefully steer the user or not.

Choice Architecture

Besides these concrete strategies to nudge decision makers, there are also some general guidelines to designing choice architectures. These strategies are mainly focused on two objectives: selecting an appropriate information load, and presenting the user with the most relevant, high-quality data [45].

According to Thaler et al. [61], there are six factors that every decision architect should take into account, regardless of their stance on nudging: incentives, mappings, defaults, feedback, error expectancy, and complex choice structure.

Incentives, as mentioned in the previous section, are a way to increase the invested effort in a decision. According to Thaler et al. [61], decision architects should always consider the user-group (who chooses, who pays, who profits, and who uses the product?) carefully, in order to select appropriate incentives.

Mappings refer to the mental concept a user has of the problem at hand. If the decision maker knows and understands the decision context and alternatives well, they are more likely to choose options that are rational. A good choice architecture should therefore aid the decision maker in their understanding of the decision domain and of the available options [61]. The system should display and explain the available alternatives in a structured way that puts each alternative into perspective within the decision context [32]. For example, if a decision maker is already aware that the standard medical procedure has a 85% survival rate, the framing of an alternative treatment (so whether the new treatment is described as having an 80% survival rate or a 20% mortality rate) is less important, because they are able to put it into perspective [45].

Defaults (see previous subsection) are one of the most prominent tools used for nudging. Thaler et al. [61] empathize that in some cases (like in the legal system), the use of defaults is unavoidable. Even a default of 'no choice or action taken' has an implicit influence on the decision maker. They propose an alternative strategy of *mandated choice* for decision architects that are more critical towards nudging. With mandated choice, it is required that the user has to select an option (e.g. otherwise the system cannot be closed/exited), but no default option is provided.

Providing decision makers with *feedback* is another option for improving decision quality. A good decision architecture should convey to the user whether they are doing well, or whether they are making mistakes [61].

Similarly, when it comes to *user errors*, decision systems should be as forgiving as possible and expect the user to make mistakes, in order to avoid frustrations [61].

Finally, the *structure of complex choices* is a vital factor when it comes to designing decision interfaces. The more options a decision problem has, and the more complex these options are, the more likely the user is to employ heuristic strategies, which tend to lead to biases [61]. Therefore, the number of alternatives should be kept reasonable, in order to avoid choice overload [32]. Besides that, the formatting, structure, and styling of the options also influences the decision maker: For example, looking at an option in an isolated way typically leads to a different treatment of the data, than when comparing two or more alternatives [45]. Another such factor that can have an influence on the perception, is the order in which the alternatives are categorized and presented to the user [32].

Besides nudging, another expression that is frequently used in discussions about choice-architecture and libertarian paternalism is the term *boost*, or *boosting* [10, 45]. The concept of boosts was introduced as a less paternalistic, and more long-term oriented alternative to nudging [29]. Boosts aim at improving the decision making process by strengthening the decision makers' cognitive and critical thinking skills, without steering them towards a certain outcome. According to Grüne-Yanoff and Hertwig [29], boosts are interventions that target competences and education (on risk-literacy, uncertainty and motivation) rather than the immediate behavior. Boosts are therefore largely identical to the previously listed non-nudging, and internal debiasing strategies discussed in Section 2.2.

Selection and Acceptance of Nudges

Similar to other debiasing strategies, the selection of a choice architecture and appropriate nudges can be a complex task. The acceptance and effectiveness of nudges is highly dependent on the user-group and their personality, and ill-chosen interventions may backfire, leading to unwanted results [32].

One major factor on the acceptance of nudges is their transparency. While some domains like marketing typically use covert nudges, decision architectures that aim for the welfare of the decision maker might consider the use of overt cues and transparent nudging instead [45]. The transparency of a nudge means whether the user is aware of the nudge's existence,

workings, and intention [10]. McKenzie et al. [45] argue that transparent cues and overt messages lessen the decision makers scepticism of a tool and strengthen their autonomy on whether or not they want to reject the nudge (which is one of the main arguments of libertarian paternalism). According to them, especially tools that try to support rationality (rational persuasion) should consider transparency, as it increases user trust [45]. Research has indicated that transparency typically does not lower the effectiveness of nudges, but that it might even strengthen it in specific cases [10].

Since nudges have been a very popular tool in the public and governmental sector, there is a lot of research on the public acceptance of nudging [10]. Generally, there seems to be a relatively high public acceptance of nudging, although it varies on the type of nudge, policy domain (health, taxes, etc.), benevolence (pro-nudger or pro-decision maker), and transparency. For example, transparent nudges that prompt reflection are seen more favourably than more covert interventions [10].

Overall, the selection and implementation of nudges is a complex topic, that should be considered carefully and adapted to the specific user-group in order to avoid unwanted side-effects or the introduction of new biases.

2.4 Organizational Dashboard Design

The term *dashboard* was originally used in the 1800s for wooden boards that sat at the front of horse-drawn carriages and stopped mud from the horses hooves from being splashed (dashed) into the carriage [30]. This term was then later adopted for the similarly placed dashboards seen in cars, which are used to quickly and efficiently communicate the vehicles status to the driver. From there on, it took hold in the business domain, describing different executive information systems [30, 21]. Nowadays, the term dashboard is mostly used to refer to systems that present and communicate visual data with the objective of analyzing, monitoring, and aiding data-driven decision making [44, 53]. Modern dashboards, in the broadest sense, can be defined as visual displays that fit the required information needed for an objective into a single screen so that it can be monitored at a glance [21].

Dashboards can have a number of unique advantages over other forms of visualizations, which should be leveraged when designing them [44]. Dashboards, compared to simpler forms of visualizations, can (1) hold more data, (2) relate more information at the same time, (3) help the user perceive the information more quickly, (4) can guide the user in reading, (5) can highlight especially important items, and (6) can be used in a way that is consistent with the company strategy [44].

Most literature typically divides dashboards into three categories according to their main purpose, namely *strategical*, *tactical*, and *operational* [5, 53, 30].

Strategical dashboards are mainly used to support decision making at an executive level, evaluating the long-term strategical goals of an organization [5, 53]. *Tactical dashboards* are used to communicate data at a mid-level (e.g. departments) to refine their tactical decision making [5, 53]. Lastly, *operational dashboards* are used to convey more detailed information about specific objectives at a lower level and shorter time-frame (often real-time or the near past) [5, 53, 30]. These categories are not mutually exclusive and can overlap in the case of more elaborate dashboards [53]. An overview of the three main dashboard categories, including their target groups, time-frames, and main purpose can be found in Figure 4. Since dashboards are becoming increasingly more popular in other domains besides business and finance, more recent research frequently mentions additional purposes of dashboards, such as education or infographics [53, 5].

Dashboard Type	OPERATIONAL DASHBOARDS	TACTICAL DASHBOARDS	STRATEGIC DASHBOARDS
Organizational Target Category	individual operations or employees	mid-level operations and departments	executive level operations and whole organization
Target Time-Frame	real-time or recent past	recent past	long-term development
Main Purpose	monitoring and optimization	analytics	high-level monitoring and steering

FIGURE 4: Overview of Dashboard Types

General Design Considerations

The main advantage of dashboards over other forms of data-display is the ability to compress large amounts of data into a comparatively smaller data space. To facilitate this, the designer must select strategies to reduce and order the amount of information shown [5]. Such a reduction or abstraction of the underlying data should be done carefully, in order to avoid misinformation or critical information loss [5].

Bach et al. [5] present a framework that summarizes the main design trade-offs to be considered in dashboard design. A simplified visualization of that can be found in Figure 5. According to them, the main four parameters that influence the effectiveness of dashboards are the available *screenspace*, *the number of pages*, *the level of abstraction*, and *the level of interaction*. These parameters influence each other, and an increase of one of them should be counterbalanced by a decrease in the others. For example, in order to show data in a less abstract way, the designer should consider increasing the number of pages, the screenspace, or the offered interactions [5]. As a general rule, Bach et al. suggest optimizing the screenspace and reducing the number of pages, in order to avoid information overload [5].

When it comes to the concrete design of dashboards, most literature divides the design considerations into two categories: (1) selecting the right content (data, meta data, visual representations), and (2) selecting the right composition (layout, design, interactions) [30, 5]. The most important considerations for each of the two categories will be discussed in the upcoming two subsections.

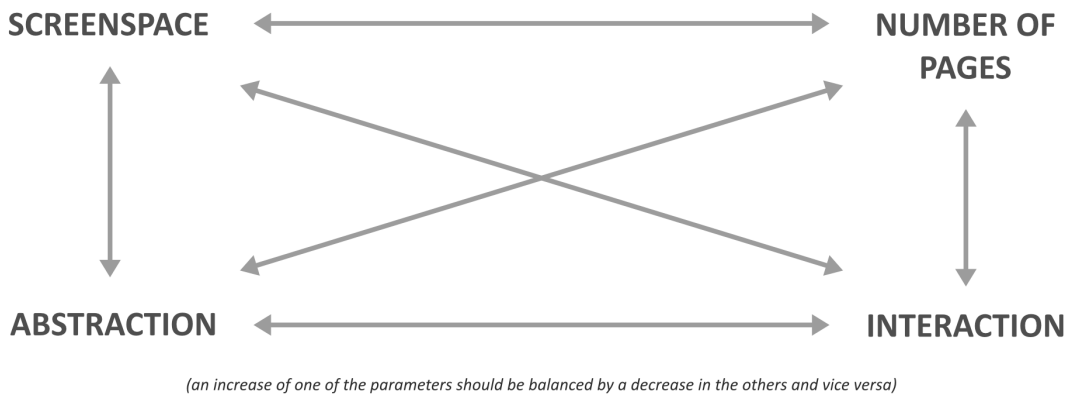

FIGURE 5: Simplified Visualization of Dashboard Design Parameters

Dashboard Content

When it comes to the content of a dashboard, the two main points are selecting the specific data, and choosing a suitable strategy on how to display it. This should be determined through a requirement analysis, focusing on the users, tasks, and goals [30, 5]. Janes et al. propose the use of elements from the GQM (Goal-Question-Measurement) models for this [30].

According to Sarikaya et al., when specifying the target audience of a dashboard, the designer should especially consider the level of circulation, the required visualization literacy, and the required domain knowledge [53].

The *level of circulation* (Public, Social, Organizational, Individual) determines how public a dashboard is: The more public a dashboard is, the less personal or sensitive data can be displayed in the visualizations [53].

The *required visualization literacy* describes the level of education and expertise needed to understand the selected visualizations [53]. For example, basic visualizations such as bar- or line-charts require a low level of visualization literacy, and can be used in most cases. Slightly more advanced visualization, such as scatterplots, heatmaps, or graphs with more axes require a medium level of visualization literacy. Finally, more advanced elements such as radar-charts, treemaps or error intervals may require a high level of visualization literacy and should only be used in contexts with an appropriate target group [53].

Similarly, the *required domain knowledge* is also an important factor to consider, especially when it comes to the level of abstraction in the data [53]. If a dashboard is only

intended for expert use, it can include more abstract and complex data than a dashboard intended for public use.

After the target audience, goals, and data for the dashboard have been determined, the question remains how that data can be communicated effectively to the user. Some of the most common elements found in dashboards are visualizations, tables, numbers, pictograms, and gauges [5]. Of these categories, visualizations are by far the most prominent in dashboards. Regarding visualizations, there are many different types that can differ in their effectiveness to communicate certain findings. In general, the most widely known chart types are bar-, line-, and pie-charts, as well as histograms [50]. This makes them relatively accessible to broad groups of users, in comparison to other charts that require a higher level of visualization literacy [53].

The different types of charts usually have different advantages and disadvantages depending on what the main purpose of the visualization is. According to Saket et al. [52] and Rodrigues et al. [50], the most common tasks that are performed with visualizations are retrieving values, making comparisons, finding extremes, finding correlations, finding clusters, finding anomalies, determining ranges, and characterizing distributions.

When *retrieving values*, bar-, line- and area-charts perform best [50]. For *making comparisons*, bar- and line-charts are the most effective [50]. In order to *find extremes*, area- and bar-charts are the most effective chart types [50]. For *finding correlations* in data, line-charts and scatterplots are the most effective, while tables and pie-charts should be avoided [50, 52]. Pie-charts and bar-charts are especially effective for finding *data clusters*, proportional relationships, and showing part-whole relations in data [52]. For *finding anomalies* in data, histograms and scatterplots perform best [50, 52]. In order to *determine ranges*, (stacked) area- and bar-charts perform best [50]. When *characterizing distributions*, histograms and boxplots are the most effective chart types [50].

A brief overview of the recommended chart types according to [50] and [52] for different user tasks can be seen in Table 2.1.

Dashboard Design and Composition

The design of a dashboard refers to the way the selected content is structured and presented to the user. These design-choices are especially important since dashboards usually compress vast amounts of data into a comparatively small space.

task	**recommended charts**	**discouraged charts**
retrieving values	bar, line, (stacked) area, scatterplot, table	bubble, stacked bar
making comparisons	bar, line	(stacked) area, bubble
finding extremes	(stacked) area, bar, line, scatterplot	bubble, stacked bar
finding correlations	line, area, bubble, scatterplot	table, pie
finding clusters	bar, pie	line
finding anomalies	scatterplot, histogram, boxplot, bar	bubble
determining ranges	stacked area, bar, line	bubble
distributions	histogram, bar, scatterplot	

TABLE 2.1: Recommended and Discouraged Chart Types per Task

Structured and visualized data is generally easier to understand than unstructured or textual data, since it makes use of the human pattern recognition abilities [44]. This makes it effective as a tool for conveying complex data. When structuring and designing a dashboard, the four main aspects are page organization, interaction, color, and typography [44, 30].

Page organization refers to the purposeful layout of content on a page. Since most dashboards (61%) only use a single page [5], the use of this limited available space should be tailored to the specific objective. According to Martins et al. [44], the layout should be structured in a way that utilizes the typical attention pattern, as seen in Figure 6. This type of diagram, also known as the Gutenberg diagram, shows a typical pattern of attention for western users (reading left to right). Typically, the upper left area is noticed first, followed by the upper right area. The attention flow then typically moves to the lower left area (as it would when reading a new paragraph) and ends at the lower right corner, completing a ‚Z'-pattern across the screen. The secondary areas (top right and bottom left) in this diagram are called fallow areas, as they typically receive less attention. Keeping this in mind, the most important information of a dashboard should be put in the upper right sector (primary optical area), and then follow the reading gravity across the screen [44]. While the lower right corner (terminal area) typically receives the least amount of attention, it is also typically the last viewed sector, making it a good place for a call-to-action element [44, 19]

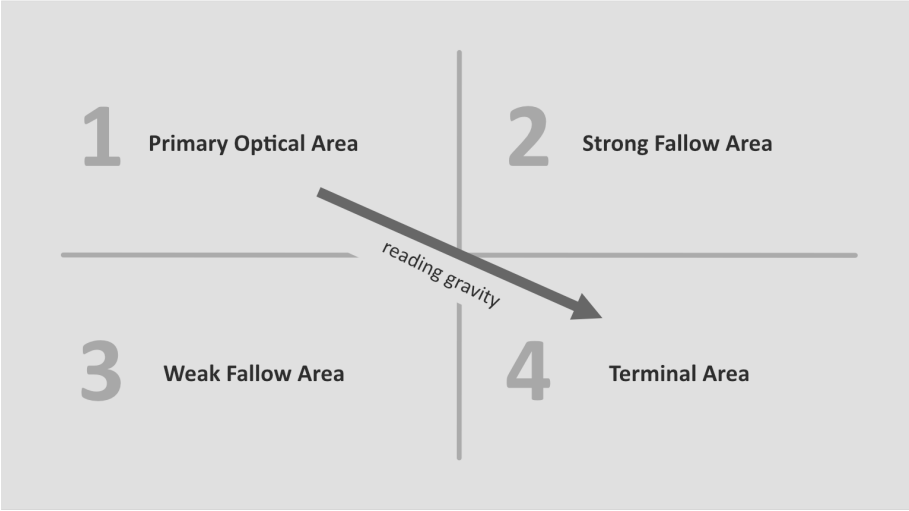

FIGURE 6: Gutenberg Diagram and Attention Direction

According to Bach et al. [5], most dashboards use this kind of stratified layout (ordered on importance), followed by semantic groups (ordered on content), and open layouts (no ordering) where users can place the widgets themselves.

When determining the page layout, the design should always include sufficient whitespace. This means the deliberate inclusion of blank spaces within the design [44]. Such whitespace is crucial to the user experience, as it helps distinguish the elements from each other and creates a visual hierarchy. A lack of whitespace can make the reading and comprehension harder, and decreases the usability [44].

Besides the page organization, the *level of interaction* is another main point to consider: in this case, the level of interaction mainly describes the way data is presented to the user. According to Janes et al. [30], this can either follow a *push* or a *pull approach*. In a *push approach*, the data is visualized in a way that guides the user through the dashboard, and provides them with information without any kind of interaction (the data is pushed towards the user). This minimizes the time needed to consult the dashboard, and typically leads to more attention [30]. Following a *pull approach*, only the most important data is presented initially, and the user can use interactions (detail on demand) to see more information. This typically de-clutters the visualization and allows for more freedom with the design-choices [30, 5]. Sarikaya et al. [53] list three more types of commonly used interactions: customization of the dashboard, faceting of the data-set, and data modification. This includes the application of filters (such as a time range or other specifications), and the option to modify the underlying data using the dashboard.

The *color* of the used visualizations is another factor that can impact the usability and attention. As a general rule, color should be used with moderation, since it can draw the users attention away from the intended path of attention [44]. Martins et al. [44] suggest designing a new dashboard completely in grey at first, and then gradually adding color to communicate certain findings. Martins et al. [44] differentiate between three different types of commonly used color-schemes: sequential (color elements according to their high- and low-values), divergent (mark exceptions and critical points in contrast color), and categorical (different groups of data are set apart through different colors). In addition to this, Bach et al. [5] mention *semantic color schemes*, for example the classic green-yellow-red color-coding usually used for traffic lights, which is often used to portray statuses on dashboards. Regardless of the selected scheme, the accessibility of the chosen colors should also be kept in mind. Users with visual impairments might have difficulty differentiating between certain color-shades, if the contrast is not high enough [30]. This can be tested by transforming the design into greyscale and seeing if the different shades can be distinguished.

The *typography* is another main factor to be considered when designing dashboards. Similarly to the color-scheme, typography can also be used to draw attention or to diminish certain elements of a dashboard through the use of different font-weights or font-types. Generally, the readability should always be the top priority when it comes to selecting the font-type and weight, and certain accents (such as boldface or italics) can be added in moderation to highlight data [44]. This is a strategy called *pre-attentive processing*, where certain important elements are highlighted (for example through color or typography) to speed up the cognitive processing of the data [30].

With these four main factors in mind, there are some general design guidelines that should be followed. First, the readability, consistency, and the visual comfort of the user should always be the top priority. This includes the fact that functional issues should always be prioritized over aesthetic ones [44]. Second, there should be a balance between visual and functional features [5]. Third, the dashboard should also not be cluttered, neither visually nor content-wise, in order to avoid overwhelming the user [5]. Finally, the graphic elements should be clearly organized and structured, with appropriate levels of interaction and complexity [5].

Challenges and Acceptance

Dashboards, like many tools that are introduced to existing workflows, can face issues when it comes to acceptance. According to Janes et al., a dashboards success is directly tied to the extent to which it follows the findings of the technology acceptance model (TAM) [30]. This refers to the perceived usefulness and the perceived ease of use of a tool or system [30].

Four main challenges that are frequently mentioned in regard to dashboard acceptance are: missing contexts, extensive volume of data, poor choices of visual design and interaction, and poor integration in business contexts [53, 2].

Preserving the *context* of data in dashboards is especially important since many visualization techniques aim at reducing the information show on screen, for example through filters or aggregations. This shapes the way that users perceive and interpret the data. Therefore visualizations are inherently not neutral, similar to the way that messages are worded or framed (see Section 2.1) [53]. In order to increase transparency and aid interpretation, the users should be informed about what was done to the data, and how it was collected [2, 53].

Another common pitfall of dashboards is the *volume of data* that is shown. If a dashboard tracks too many metrics and indicators, the user can become overwhelmed and get sidetracked from the important parameters [2].

This often is combined with *poor visual design and misleading interactions* [2]. Research has shown that even simple interactions can prove too complex for some user groups, and that in most cases some kind of guidance or explanation is needed [2].

Finally, a common criticism of strategic business dashboards is a *lack of integration* and *poor linkage with the actual business strategy* of a company. Managers often remark that there is a disconnect between the dashboard content and the actual business strategy, especially in cases where the dashboard has been imposed in a top-down manner [2].

These four critical factors should be kept in mind when developing dashboards, in order to improve user acceptance. Allio [2] proposes the use of management dashboard in combination with other business tools, such as automatic review and analysis tools, as a way to enrich the context, improve the strategic alignment, and increase the user acceptance.

Chapter 3

The Proposed Dashboard and Decision Support System

This chapter will describe the proposed system in detail, including the requirements and the selected technologies for each component. First, the requirements analysis will be discussed, covering the project scope as well as both functional and non-functional requirements. Next, the selected technologies will be presented, along with the rationale behind each choice for the system components.

3.1 Requirement Analysis

3.1.1 Project Scope and Objectives

The main objective of the project is to build a tool that strengthens rational decision making and supports project managers in their decision making process in regard to risk-related cognitive biases.

As mentioned in Sections 2.2 and 2.3, the key to debiasing irrational decisions is creating bias awareness and a motivation to correct those biases. According to Bhandari et al. [6], financial decision makers should learn from their past choices and outcomes and carefully monitor the present situation, in order to make more informed choices for future decisions. In order to facilitate this, the proposed tool needs at least two parts: a feedback step that informs the user about possible biases while making new decisions, and an overview of past decisions and their outcome and impact on the business. This is done using a dashboard and a decision support screen that can be integrated into existing project management

software. Since dashboards are a frequently used tool in business contexts for monitoring data over time and conveying large amounts of data in a relatively compact format, the use of a dashboard is suitable to present the user with past decision data and outcomes, as well as information on biases in general. This is then extended through the use of a decision support screen to support the user when making new choices.

To limit the scope of the project, it is assumed that previous and current biases are automatically identified and stored in a database. Common approaches for the identification of biases in business and project management include the use of machine learning [49, 51], formal decision models [26, 16, 51], or multi agent systems [26, 51].

This thesis focuses on how to effectively communicate this data to the decision maker, and how to present past decisions and their outcomes in a way that helps the decision maker make more informed decisions. Due to the added complexity of multi-criteria decision making, this research will only take into account go/kill decisions with two options: whether to continue or stop a project.

3.1.2 Functional Requirements

The main stakeholders for this project are project managers that want to keep an overview of their past decisions and improve their decision making process. In line with the project objectives, the main use cases for this tool are: (1) A project manager wants to see their decision history and reflect on it, (2) A project manager wants to see a specific past decision, and reflect on the outcome and possible biases, (3) A project manager wants to make a new go/kill decision. These three use cases include reflection on past decisions and their outcomes (past and present), as well as supporting informed decision making in future scenarios.

Based on this, the following user stories were created for the three previously mentioned use cases. A use case diagram for the system can be seen in Figure 7.

> (1) *As a project manager, I want to be able to see my decision history in a dashboard, so that I can evaluate the impact on the business and detect emerging trends in the data.*

> (2) *As a project manager, I want to be able to look up specific past decisions, so that I can reflect on their outcome and check for possible improvements.*

(3) *As a project manager, I want to be able to make well-informed go/kill decisions, so that I can minimize biases and financial risks.*

The main functional requirements for the system that can be derived from these use cases are listed below.

Provide a decision history overview (dashboard)

1. show decision history over time: The system should display an overview of past decisions and have the option to look at specific decisions in detail (including potential biases and outcomes).
2. show financial impact of decisions: The system should provide a overview of the financial impact of past decisions, including losses and gains associated with the decisions.
3. show outcome of decisions: the system should show an overview of the outcomes of decisions, for example whether a project failed or succeeded.
4. show information related to potential biases: The system should provide information on potential biases and give a reasoning for the classification of past decisions.

Provide a decision support screen (for new decisions)

1. provide relevant decision data: The system should summarize the most important data needed to make a go/kill decision in an easily comprehensible format
2. provide interface to select option/make decision: The system should provide the interface needed to make and finalize a new go/kill decision
3. provide feedback about decision: the system should be able to provide the user with feedback on the decision and warn about potential biases they might be susceptible to.
4. support modifying the decision: The system should allow the user to modify their decision after receiving feedback.

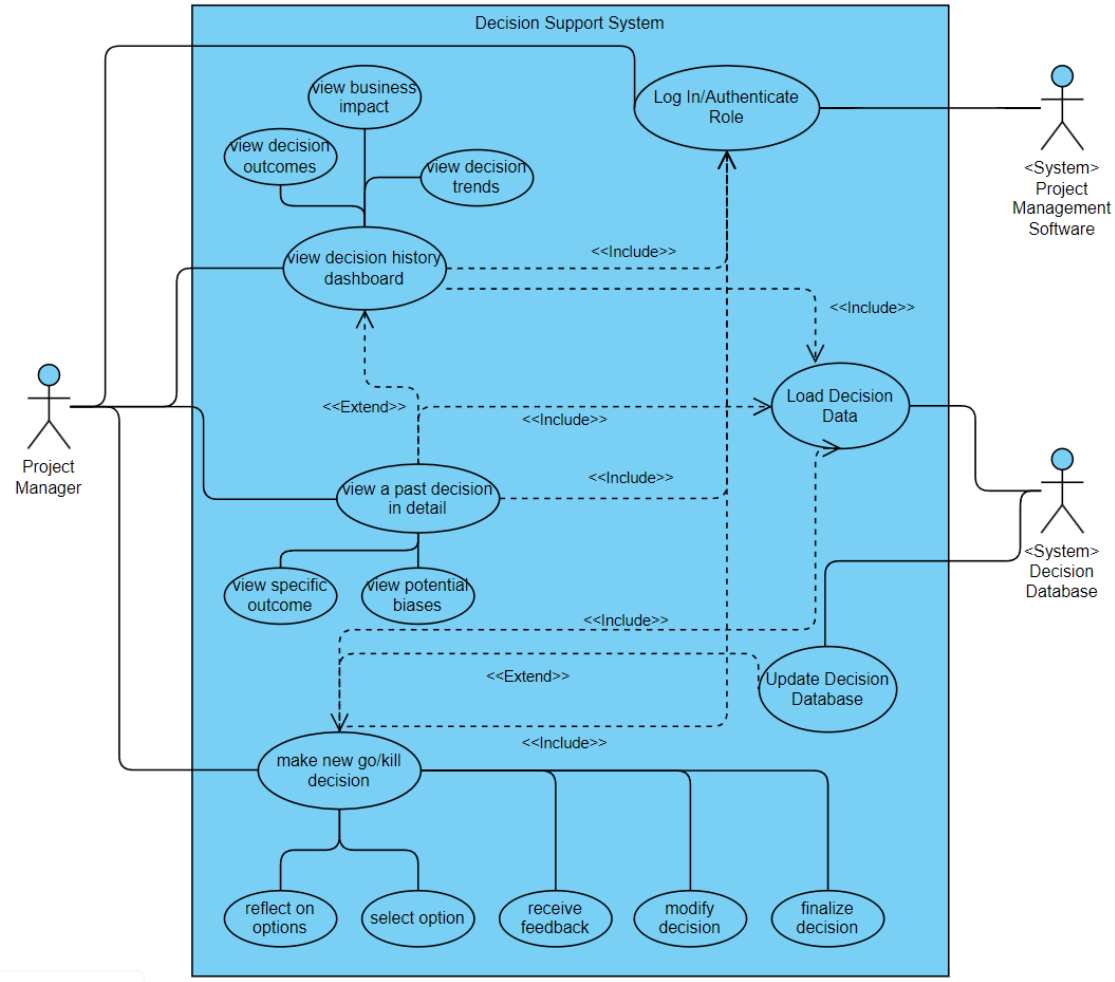

FIGURE 7: Use Case Diagram

3.1.3 Non-functional Requirements

The main non-functional requirements for the system can be derived from the literature findings on the user acceptance of debiasing strategies, nudging, and dashboard design.

As mentioned in Section 2.4, the key points for user acceptance in dashboard design are tied to the technology acceptance model (TAM), consisting of the perceived usefulness and the ease of use of a system [30]. Similarly, the volume of data, the business integration, accessibility, and the visual design itself also play a key role in the user acceptance of such systems [53, 2, 30]. The content and design should be tailored to the specific target group, in this case project managers as the main stakeholders.

Besides this, a number of general requirements, such as reliability and availability should also be considered. Based on these considerations, a number of non-functional requirements can be derived:

Non-functional Requirements

1. Ease of use/usability: the system should be intuitive and easy to use for the specified target group (project managers), and follow common usability guidelines.
2. Usefulness: the system should provide relevant information and offer new insights about cognitive biases for project managers wanting to improve their decision making.
3. Data: the system should use appropriate data and metrics, with which the target group is familiar. Furthermore, the system should be able to handle and summarize large amounts of data into a more accessible format for the user. The displayed metrics and data should be accurate, consistent, and tailored to the user group.
4. Design: the design of the system should adhere to common usability standards and accessibility guidelines. It should be professional and appeal to project managers as the main stakeholders. Furthermore, it should support rational decision making through the use of nudges.
5. Business integration: the system should be easy to integrate into existing project management software and workflows, in order to increase the user acceptance.
6. Robustness: the system should be reliable and robust to user errors.

3.1.4 System Architecture

The design for the system architecture is based on the previously described requirements. The system consists of two sub-systems: a dashboard for displaying and analyzing historic decision data, and a decision support system (DSS) to support real-time decision making. These two components are implemented on top of an existing project management tool (requirement 5), in order to allow for easier integration into existing workflows. The decision data and the information on potential biases are stored in a relational database. This database is partly based on existing bias ontologies [54]. The details of the implementation are described in the following two sections. A schematic overview of the system can be seen in Figure 8.

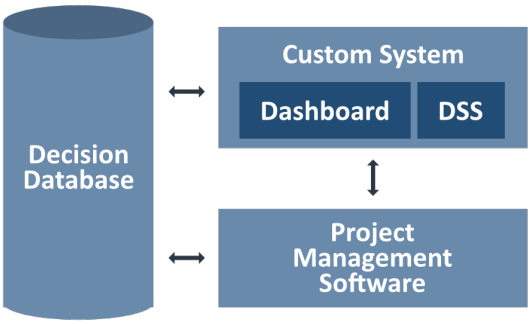

FIGURE 8: Schematic Overview

As for the system functionalities, an overview of the use cases per component can be seen in the activity diagrams in Figures 9 and 10.

For the dashboard (Figure 10), the default use case starts with the user opening the dashboard. The system then loads the data from the database and updates the interface accordingly. The user can then use the dashboard, by reflecting on past decisions, analyzing the data, or looking for specific trends and developments in the visualizations.

The default workflow for the decision support screen (the screen used when making new go/kill decisions) is the user opening the support screen, followed by the system loading, calculating, and displaying the relevant data (Figure 9). The user can read and evaluate the data before making a selection. The user will then have to provide a reasoning for the choice before submitting it. The system will then evaluate the data. If the user picked the rational option, the system will display a confirmation message and save the data, so it can be displayed in the dashboard or other user interfaces. If the user picked the risky option, the system will issue a warning, and provide the option to modify the decision. The user can either accept or reject this. If the user rejects the modification offer, the system will send a confirmation message and save the data. If the user wants to modify their choice, they will return to the decision making screen. The saved decision data is then saved to the database and displayed in the dashboard the next time the user opens it.

3.2 Technologies Used for Implementation

Regarding the choice of technologies, the main focus was to keep the selection open-source and free. This includes the project management software that is used as the basis for the implementation, any additional libraries and frameworks used, and the proposed database system.

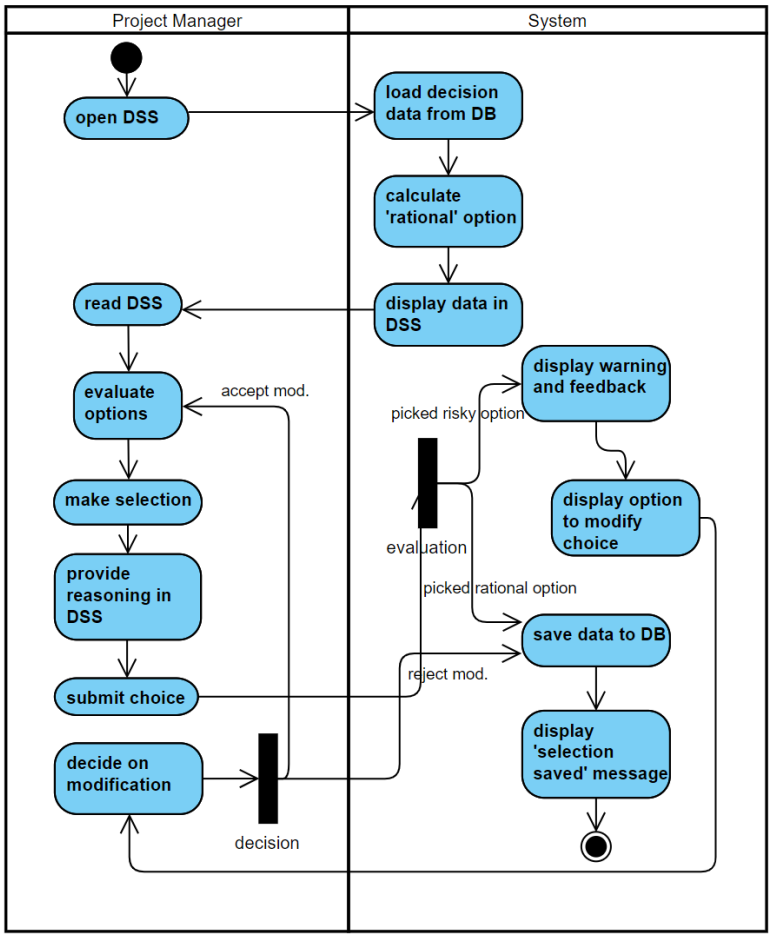

FIGURE 9: Activity Diagram DSS

Project Management Software

A number of free project management software options were considered and compared for this project, including monday.com[1], ClickUp, Teamwork and Wrike [25]. From these four options, monday.com was found to be the most suitable due to their extensive support for third-party development, including an API, a software development kit (SDK), and a design framework. This allows for a range of different third-party integrations, such as custom widgets, dashboards, pages, and automations.

As for the underlying technologies, monday.com mainly relies on GraphQL for internal communication and storage. For third-party app development, most web-based technologies can be used (including plain HMTL, JavaScript and CSS), but it is recommended to use React as the main JavaScript framework in combination with Node.js for running and delivering the app [47].

[1] `https://monday.com`

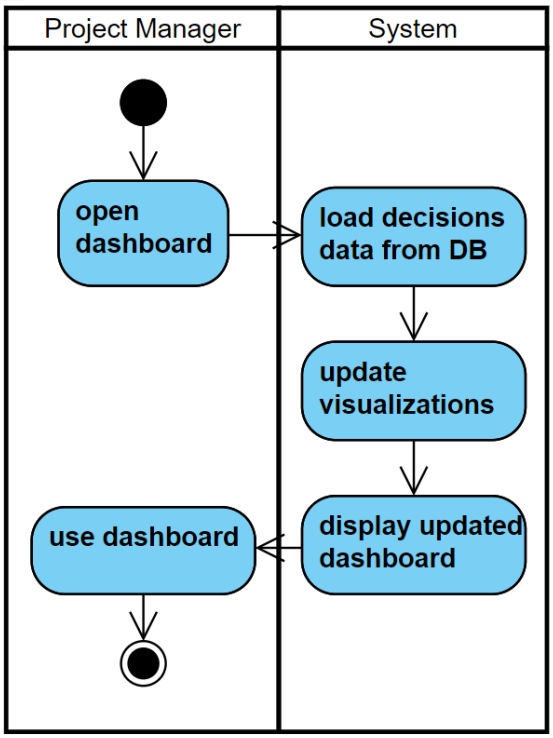

FIGURE 10: Activity Diagram Dashboard

In addition to these guidelines, monday.com also offers a UI/UX design framework with best practices and additional design components that can be used in custom apps [48].

Visualization Software

Besides the default React and JavaScript functionalities, and additional JavaScript framework called Recharts[2] was used for visualizations. This is a free and open-source React framework for basic data-visualizations, such as line- or pie-charts.

Database

Because the internal storage functions of monday.com are not sufficient to map complex relationships, an external database is required. For this, it was agreed on to use a relational database system to allow for easier collaboration and exchange. In this case, the preferred software was PostgreSQL, which is free and open-source[3]. A visual overview of the used technology stack can be seen in Figure 11.

[2] `https://recharts.org/`
[3] `https://www.postgresql.org`

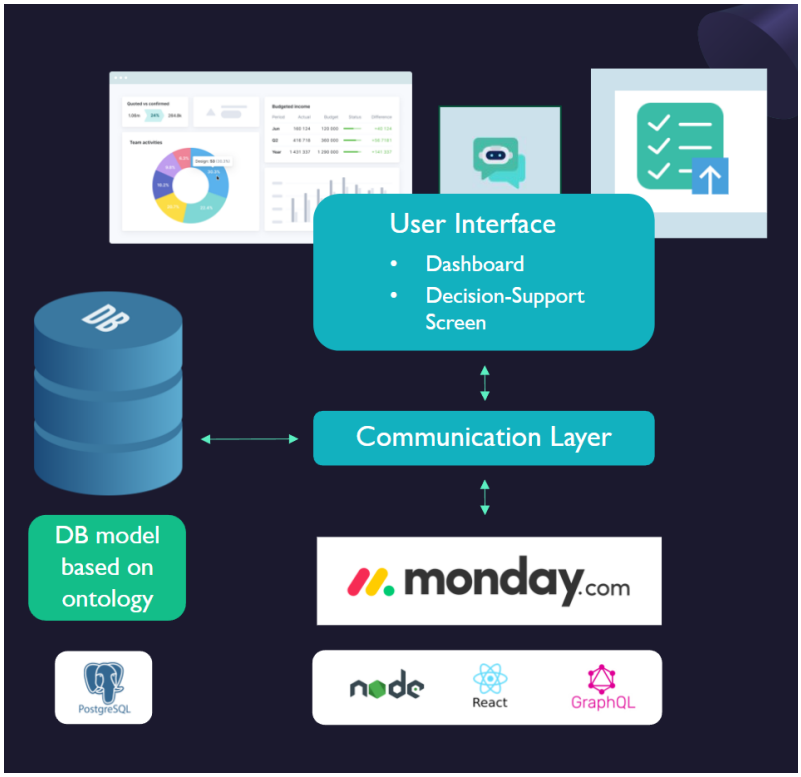

FIGURE 11: Technology Overview System Architecture

Chapter 4

System Design

The design and content of the system is based on the design principles found in the literature discussed in Chapter 2. First, the considerations and design choices for the dashboard will be discussed. Then, the design choices for the decision support screen are described, followed by the proposed database structure.

4.1 Dashboard Design

According to the typical dashboard classification of *strategical*, *tactical*, and *operational*, the proposed system mainly falls into the categories of operational and tactical [5, 53, 30]. The dashboard is intended for individual use, therefore falling into the operational category regarding the scope level, and its main purpose is analytics and the review of past decisions, therefore also falling into the tactical category.

4.1.1 Content

The main goal of the dashboard is to inform the user about their past decision history and about cognitive biases and their impact, as defined in the requirement analysis (Section 3.1). This is meant to create bias awareness and a motivation to correct biased behaviour, which is required for successful debiasing [14]. Overall, the dashboard content is primarily based on internal debiasing strategies (modifying the decision maker), namely *training in biases*. This strategy aims at educating the user about cognitive biases, how to recognize them, and how to avoid them [40].

The proposed dashboard should therefore contain the following data and information:

- Decision data over time
- Detailed decision data
- Impact and outcome of biased decisions
- Information/feedback on own behaviour
- Additional information on cognitive biases

According to the classification by Sarikaya et al. [53], the level of circulation of a dashboard determines what type of data should be displayed. Since this system is mainly intended for personal analytics and not for public use, the dashboard can contain sensitive or confidential business data.

Regarding the required level of visualization literacy [53], at least a medium level of literacy can be assumed since the target group consists of management professionals with a financial background. This allows for the use of more complex graphs and other visualizations, such as scatterplots and heatmaps [53]. Any visualizations that require higher levels of visualization literacy are not used, in order to increase the usability and accessibility of the dashboard.

Similarly, the required domain knowledge can be assumed as medium to high, since the dashboard is not indented for public use, but solely for professionals. The dashboard can therefore contain more abstract data, since the target group is professional users [53]. This should be kept in mind for the validation, since testing the system with non-professionals could limit the effectiveness of the chosen visualizations.

As for the individual visualization charts used in the dashboard, the proposed design is mainly based on the recommendations by Bach et al. [5], Rodrigues et al. [50], and Saket et al. [52]. Since the most widely used visualization types are tables, numbers, pictograms, and gauges [5], as well as bar-, line-, pie-charts, and histograms [50], this makes them relatively accessible and easy to understand to most users.

According to Rodrigues et al. [50], area-, and line-charts perform best for retrieving values and finding extremes in data, therefore making it easier to understand trends and developments. Based on this, an area/line-chart is used in the dashboard to show the development of decisions over time in a timeline component.

To extend the information from the timeline item, and to allow for a detailed review of past decisions, a table is used to show a detailed view of the past decisions. Similarly, to provide more context and information about the mentioned biases, an information-box with additional explanations, examples, and strategies to avoid the biases is added to the dashboard.

Pie- and bar-charts are especially effective for determining proportional relationships and part-or relationships [52]. The proposed dashboard therefore uses pie-charts to convey any data that appears in part-of relationships, such as the outcome of decisions, the amount of decisions that were debiased (or reconsidered), and additional metadata about the decisions. To differentiate between the different pie-chart components, a combination of full- and half-pies is used. For any metrics that can be summarized in a single number or average, number elements are used to display this information in the dashboard.

Since preserving the context and origin of the used data is very important for user acceptance [53, 2], the dashboard provides additional information about the data in form of context clues and information boxes.

The content of the dashboard with the used visualization types can be summarized as follows:

1. Area/Line-Chart: overview of (biased) decisions over time
2. Table: detailed information about individual decisions
3. Text/Info-Box: detailed information about cognitive biases
4. Pie-Charts: data about biased decisions, outcomes, decision process (part-of-whole relationships)
5. Numbers: average metrics about biased decisions and their outcome

4.1.2 Layout and Design

According to Bach et al. [5], the main four design trade-offs that should be considered for dashboard layout are screenspace, the number of pages, the level of abstraction, and the level of interaction. Most typical dashboards in business settings use a single page with optimized screenspace [5]. To adhere to this common design principle for business dashboards, the screenspace and number of pages are also kept low in this system (one-page approach), which means an increase in abstraction and interaction is needed to balance

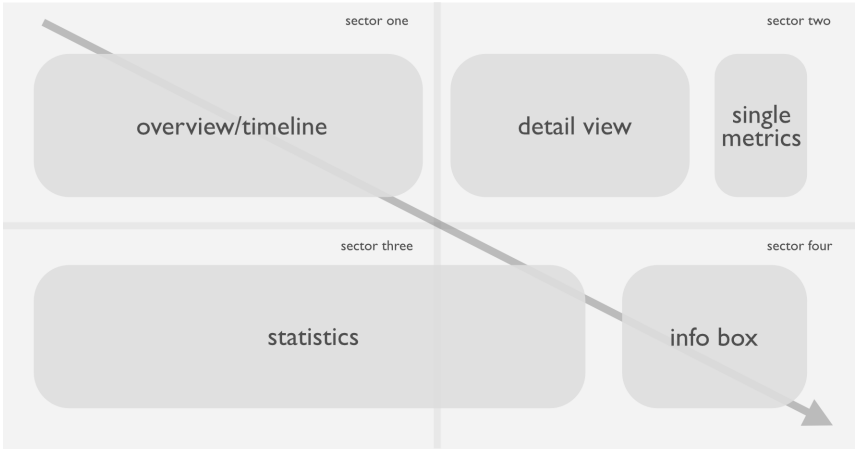

FIGURE 12: Screen Layout

those factors out. As mentioned in the previous section, the target group should be able to handle these higher levels of abstraction and interaction, which allows for more complex designs.

For the layout of the content on the page itself, a stratified layout (ordered on importance) according to the Gutenberg attention pattern is used in order to maximize the effectiveness of the visualizations [5, 44]. Following these guidelines, the most important visualizations (the timeline and the detailed decision data) are placed in the primary optical zone (top left), while the less important visualizations (complimentary visualizations (pie charts) and the average metrics) are placed in the fallow areas (top right and bottom left). Finally, any information that should be retained, or that gives a call-to-action should be placed in the terminal area of the layout (bottom right). In this case, the terminal area is most suitable for the information box on biases, since it contains information on how to detect and avoid cognitive biases. A rough layout of the items can be seen in Figure 12, and the detailed dashboard design can seen in Figure 13.

To structure this design into a coherent layout, and to add sufficient whitespace, this design was implemented using a grid system. Vertically, there are two sections (top and bottom half) to create a structure similar to the Gutenberg diagram. Horizontally, the design is divided into nine sections, to allow for a detailed layout. In-between the sections a gap is added to create whitespace and improve the readability of the design. A draft of the used grid system can be seen in Figure 14.

Regarding the level of required interaction (push or pull approaches [30]), this system is best suited for a push approach, where the most important data is pushed towards the

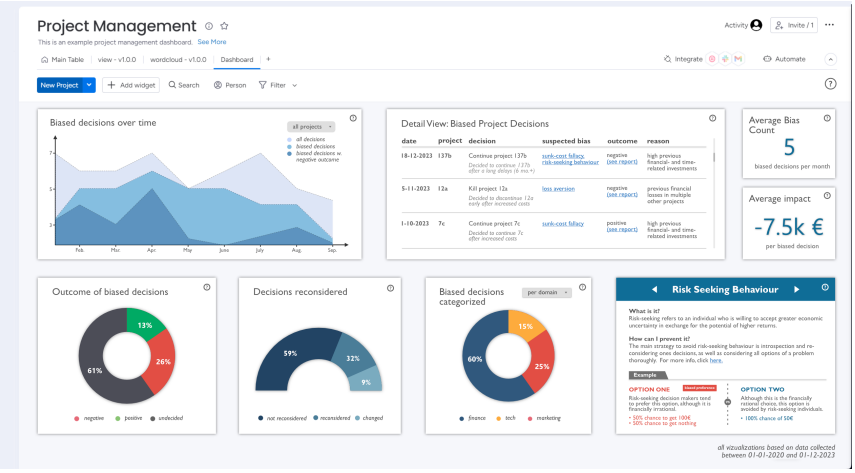

FIGURE 13: First Dashboard Design

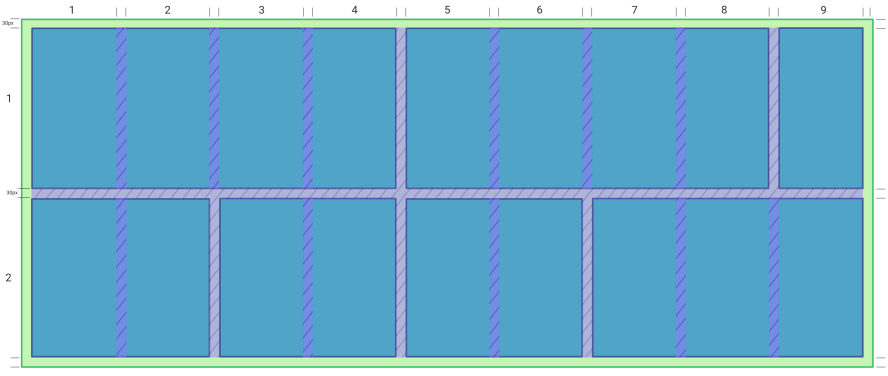

FIGURE 14: Grid Layout

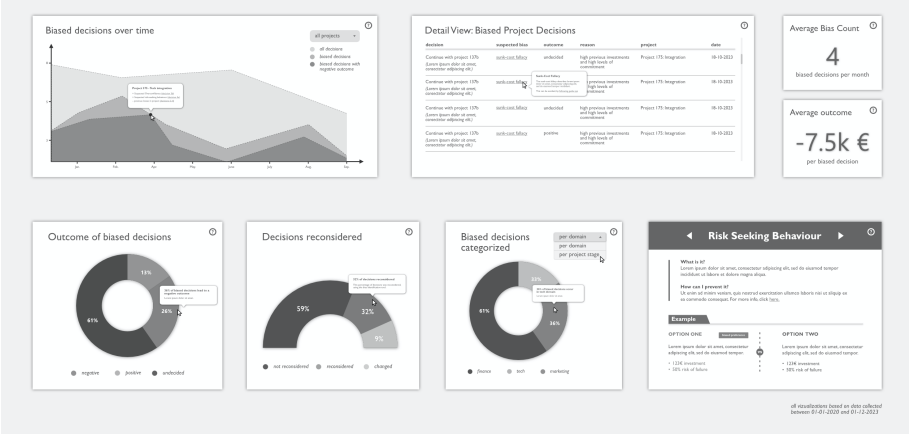

FIGURE 15: Dashboard in Greyscale

user and accessible without any interaction. If the user requires more detailed information, this is offered through a pull approach, where the user can see detail on-demand (for example through the use of hover interactions, filters or modal pop-ups). The detailed information includes information about the underlying data and visualizations (info-buttons), explanations (tooltips), and filtered data (menu to switch between different pie-charts).

As for the color scheme, a categorical scheme (different color schemes for different visualizations) [44] seems best suited, in order to differentiate between the different sections of the dashboard. In addition to this, semantic colors are added for the visualizations where there is a clear theme, such as the outcome of decisions (e.g. positive/negative outcome and green/red) [5]. Other than that, color is used sparingly and with a focus on maintaining a high enough contrast, in order to increase the accessibility for users with visual impairments [30]. A picture of the design in greyscale for contrast comparison can be seen in Figure 15.

Regarding typography, readability and accessibility are the main priority [44]. Therefore, modern and relatively unobtrusive sans-serif fonts such as Roboto or Gill Sans are best suited. These font-families offer a wide range of font-weights and styles, and are easy to read. For the digital designs, Gill Sans was used as the main font, while Roboto was used for the implemented web-version, because it is more widely available across different devices. An example of the used fonts can be seen in Figure 16. As suggested by Janes et al. [30], the design also makes use of pre-attentive processing (color, italics and boldface) in order to make the data more easily readable and to guide the user focus to the most important bits of data. This can be seen in the designs and font-choices of the data tooltips, the bias information box, and the average metrics.

FIGURE 16: Selected Fonts

4.2 Decision Support Design

The main purpose of the decision support screen is to support the decision maker in making rational and well reasoned choices. Therefore, the screen should summarize the most important data in a compelling and easily understandable format for the decision maker. Based on this, and the requirement analysis (Section 3.1), the support screen should at least include the following data:

1. Project description and context
2. Current status of project
3. Reason for re-evaluation
4. Go/Kill options with relevant data summarized (like risk estimation, financial status)

In order to complement the dashboard, which aims at educating the user and creating awareness through internal debiasing, the decision support screen mainly relies on external strategies and nudging for debiasing. An overview of the selected debiasing strategies and reasoning whether they are suitable for the dashboard and decision support screen can be found in Table 4.1.

Overall, the entire decision support screen falls into the category of external technological debiasing, since it acts as a computer-aided decision support [57]. This is supported through the use of other external debiasing strategies and nudges, as well as a structured choice architecture, which provides the user with the most important data in a easily comprehensible format.

As for the external debiasing strategies, monetary incentives and social factors are not suitable for this specific use case. Social factors are outside the scope of the system since the decisions require expert and domain knowledge and may contain confidential company data. Therefore sharing the choices or comparing them to a public consensus makes little

sense. Similarly, monetary incentives fall outside the scope of this project, since the decision making takes place in a corporate work-setting.

Compared to that, accountability as a more social form of incentive is better suited for this specific use case. If there are higher (social) stakes related to the decisions (social accountability), decision makers tend to invest more effort into the reasoning [57]. The decision support screen therefore provides a text box which requires the user to give a short explanation and their reasoning behind their choice for the record. This gives the impression that the user might have to justify their decision to others, which increases the social stakes [57].

Another option to apply external debiasing is the use of framing [32]. In this case, the goal is to present the data in the most accessible way, using framing and highlighting of certain elements, as well as displaying values in an easily understandable format (e.g. diagrams or frequencies). In addition to this, specific framing is used to avoid triggering terms like 'loss'.

In addition to these external debiasing strategies, more direct nudges are used to steer the user towards rational choice. Generally, the proposed nudges fall into the category of benevolent nudges, since they aim at increasing rationality and primarily benefit the decision maker (and potentially the company) [10]. Furthermore, they adhere to the guidelines set by Thaler and Sunstein [60], by not restricting any choices and being easily avoidable. In addition to this, the system will be designed with information cues, making the nudges more transparent. This is supposed to make the nudges less paternalistic and preserve user trust [45]. In addition to this, it also strengthens the user autonomy, by allowing the user to consciously reject nudges, therefore aligning with the principles of Thaler and Sunstein [60, 45].

From the four types of nudges (defaults, reflection, future-focused thinking, and information shaping) categorized by Soll et al. [57], all categories except future-focused thinking are applicable to this scenario. The nudges that fall under the future-focused thinking category all require flexibility when it comes to scheduling, which is not necessarily given in corporate contexts. Besides that, the nudges that shape information are largely identical to framing, which was discussed under the external debiasing strategies.

Defaults are considered one of the most effective nudging strategies in general, since they rely on the status-quo and the users' inertia. According to Soll et al. [57], smart defaults (that pre-select a mathematically chosen option) are especially useful for financial

debiasing. For this reason, the decision support screen pre-selects the rational option as the default, and presents it as the first option.

Reflection inducing nudges, such as planning prompts or active choice, try to steer the user towards reflection by adding an interruption to the decision making process [57]. In this case, a reflection checkbox is added to the system, and the previously mentioned accountability prompt adds a brief interruption to the process. This is meant to make the user reflect on their choices at least briefly before submitting them.

Regarding the decision architecture, the main point is selecting an appropriate information load and presenting the decision options in an easily understandable format [45].

In order to provide the user with the decision context and project information, a short textual summary is added to the top of the decision screen, with the most important data highlighted visually (boldface). In addition to this, a visual element summarizing the key points at a glance is added below the summary, in order to provide a more visual recap of the key points. Since visualized data is often easier to understand than purely textual data [44], additional miniature visualizations are added to summarize the financial and time-related status of the project.

As for the two decision options (go or kill the project), the structure, formatting, and mapping of the alternatives are especially important factors to consider [32]. Ideally, the system should display and explain the alternatives in a structured way, and put the alternatives into context [32]. Therefore, the system presents the two options next to each other, in a structured way, summarizing the key points of each alternative. This is used in combination with the previously mentioned default nudge, by pre-selecting the rational option, highlighting it visually (through color), and presenting it on the left-hand side where people tend to read it first, according to the western reading direction.

Finally, the last feature that is added as a nudge and choice architecture tool is feedback [61]. In order to debias the user and support reflection, the system will provide the user with detailed feedback in case of selecting the biased option. When the user rejects the rational default option, a pop-up with a warning about potential biases and an explanation is presented to the user, with the option to reconsider their choice or submit it. This is meant to increase the bias awareness of the user and give them the chance to reflect on their choice. Like the other nudges, this step is also rejectable, meaning the user can choose to submit their choice without any reconsideration.

Category	**Strategy**	**Is suitable?**	**Reasoning**
Internal	Training in rules	No	outside of scope
	Training in biases	Yes	inform user about biases in their own decision making and when they occur
	Training in representations	No	framed numerical values in accessible format (visualizations and probabilities) instead
External	Incentives	No	monetary incentives not applicable, since it is in a corporate/business setting
	Accountability	Yes	make user give reasoning for choice
	Social factors	No	social norms are not applicable, since expert/domain knowledge is required
	Framing	Yes	decision options can be framed with certain outcome in mind
	Technology	Yes	system acts as a decision support system and visual decision aid
Nudging	Defaults	Yes	smart defaults as suggestion
	Reflection inducing nudges	Yes	active choice and reflection prompts specifically
	Nudges toward future focused thinking	No	time-frame of decision cannot be changed
	Nudges that kindly shape information	Yes	see framing
Choice architecture	Mappings	Yes	decision support system should structure data well and provide context
	Mandated choice	Yes	could be used as alternative to default
	Feedback	Yes	give user feedback on potential biases
	Choice Structure	Yes	structure the decision data that is presented to user

TABLE 4.1: Debiasing Choices for the DSS and Dashboard

The elements of the decisions support screen are structured top-down, ordered according to the typical reading gravity and attention pattern. The project description and decision context are placed at the top, to provide a first overview of the situation. Below that, the visualizations (miniature charts) and the project at a glance section are placed, in order to summarize the most important data in a visual form. This should provide the user with all the data needed to make an informed choice. Below that, the two decision options (go/kill) are presented, with the rational choice being highlighted and pre-selected. For each option, the key points are summarized for easier decision making. Below the options, the reflection inducing elements are placed, namely the text field to provide the reasoning behind the choice, and the checkbox asking the user whether they made an informed and thoroughly reasoned choice. At the bottom of the screen, the submit and save buttons are placed. In addition to these elements, information cues are placed within the decision support screen, in order to create transparency and provide the user with some context as to where the data comes from and what the purpose of the tool is.

The sketches and design of the decision support screen can be seen in Figures 17 and 18. The design of the feedback overlay can be seen in Figure 19.

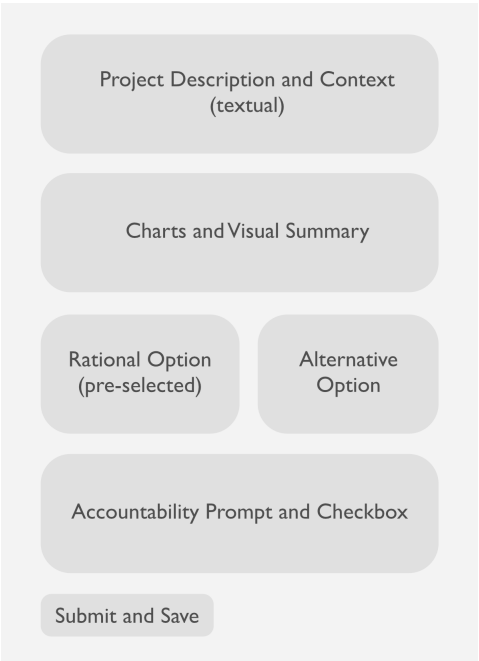

FIGURE 17: DSS Layout

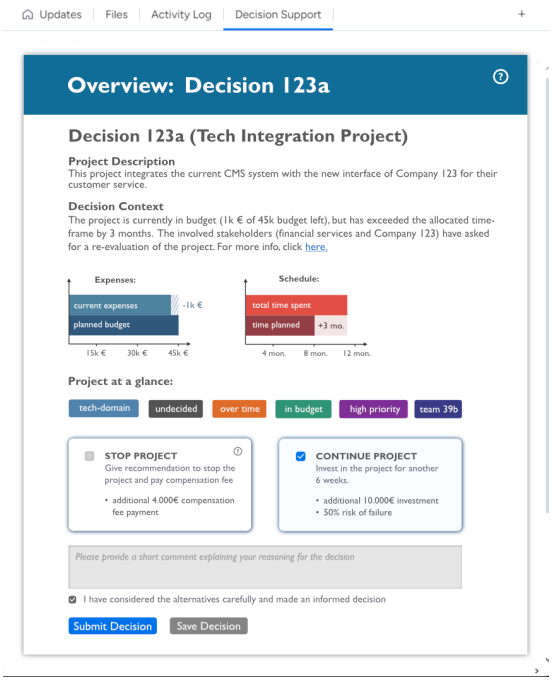

FIGURE 18: First DSS Design

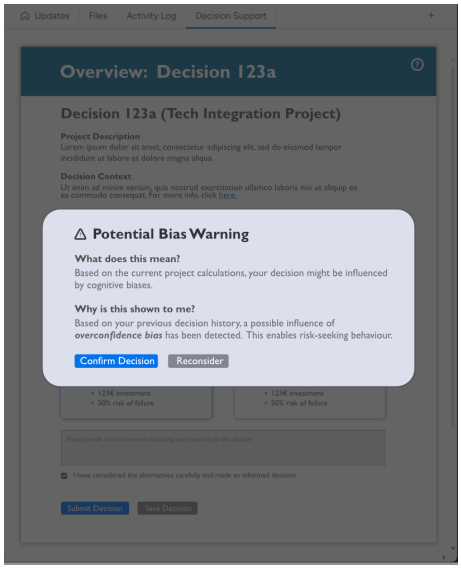

FIGURE 19: DSS Feedback Design

4.3 Database Design

Based on the design and the requirement analysis, a database design was made using Postgres, which can be seen in Figure 20.

The main entity are projects, which have a unique ID, as well as additional properties such as start- and end-date, priority, and assigned staff. Each project can have multiple related decisions, which in turn can have potential biases and positive or negative outcomes (financial- or time-related).

Due to time constraints and the need for further collaboration with other project members in order to create a consistent database scheme across all project parts, this database schema was only conceptualized and not implemented. Therefore, and in order to make user testing more easy and lightweight in terms of required software, the data was managed locally instead.

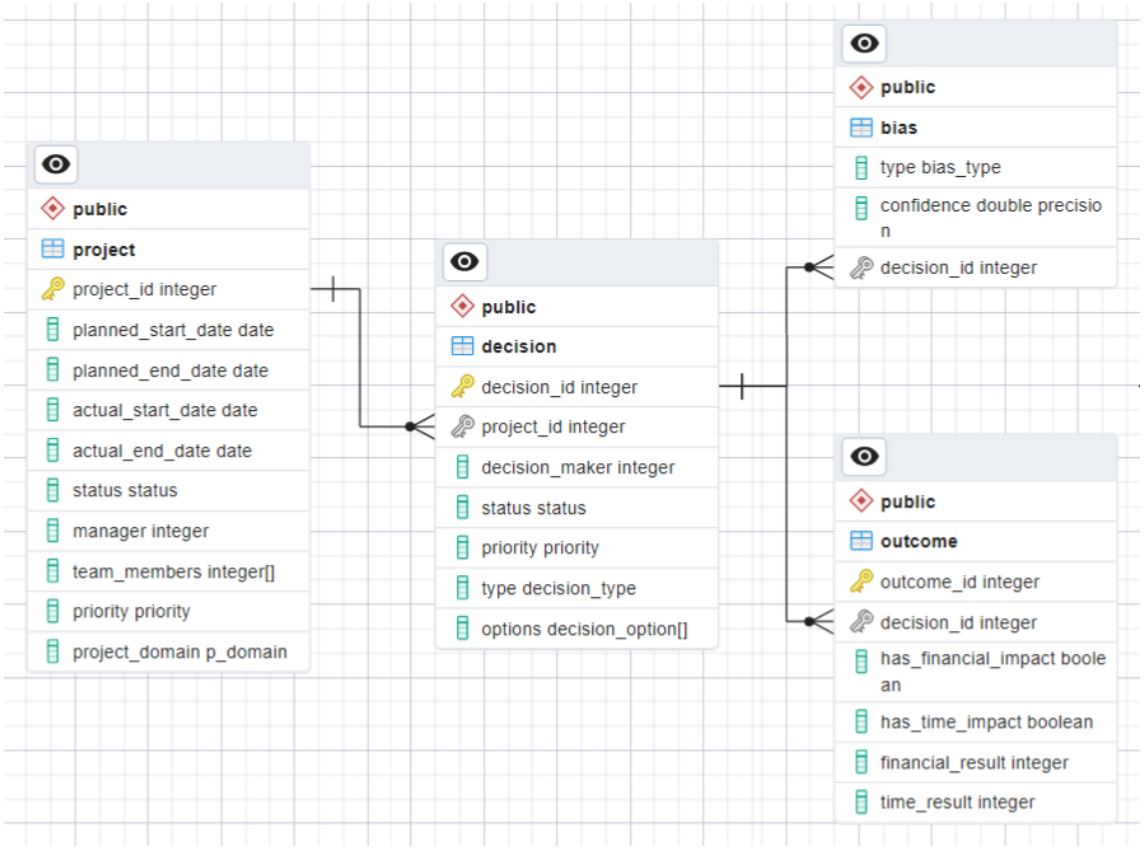

FIGURE 20: Database Schema Draft

Chapter 5

Usability Testing

Before the final implementation and validation of the system, usability tests were conducted in order to ensure the usability for the final experiment. The research design, gathered feedback, and the resulting design improvements will be discussed in the following sections.

5.1 Procedure

The procedure for the usability tests followed a 3-part structure: (1) a free exploration using think aloud, (2) task-based scenarios, (3) follow-up interviews.

The free exploration was used to let the participants familiarize themselves with the system and to gain an impression of which items particularly stand out to new users. In addition to this, the think aloud data also gives some insight about the actual path of attention that the participants follow, and whether it matches the expected Gutenberg pattern.

During the task-based part of the test, the participants were asked to first complete a number of simple tasks with the dashboard and then make a decision using the decision support screen while thinking aloud. The provided tasks required the participants to access information about past biases and their impact, about the current status of decision making, and about trends that can be seen in the data. The tasks therefore cover all three temporal dimensions (past, present, future) as mentioned by Bhandari et al. [6]. In addition to the qualitative think aloud data, additional metrics, such as the number of errors, the time spent on task, and the completion rate per user were also collected.

Dashboard Tasks:

1. Can you tell me how many biased decisions were taken in April of 2023? And can you tell me how many of those had a negative outcome?
2. Can you tell me a reason, why the last decision was flagged as biased?
3. Can you tell me a way to prevent risk seeking behavior?
4. Can you tell me what percentage of biased decisions have been reconsidered?
5. Can you tell me the average financial impact of a biased decision?
6. Can you tell me the domain, in which most biased decisions happen?
7. Can you tell me the project stage, in which most biased decisions happen?
8. Can you tell me what percent of biased decisions have a negative outcome?

As for the decision screen, the users were asked to go over the data using think aloud and make a decision using the system.

Finally, a follow-up interview on the usability and usefulness of the system was conducted. According to Almasi et al. [3], the most widely used questionnaires for assessing dashboard usability are the System Usability Scale (SUS), Technology Acceptance Model (TAM), Situation Awareness Rating Technique (SART), Questionnaire for User Interaction Satisfaction (QUIS), and the Unified Theory of Acceptance and Use of Technology (UTAUT). From these questionnaires the SUS [8] appears to be the most suitable, due to being more lightweight and general than the other questionnaires. To extend the questions from the SUS questionnaire, some questions from the PSSUQ (Post-Study System Usability Questionnaire) [42] were added. The PSSUQ questionnaire is commonly used to assess the usability, usefulness, and information of a system following a user study.

Since the goal of the interview was the collection of qualitative data about the usability of the tool, a number of custom questions were added, and the SUS and PSSUQ questions were adapted to make them more open ended and qualitative. The used interview questions can be found in Figure 21. For comparison, the SUS and PSSUQ questionnaires can be found in the Appendix, Figure 38.

The target group for this usability research was people above the age of 18 (able to consent) who are either interested in, or have experience with project management

Adapted interview questions

Usability

- Issues with design, structure?

1. What was your overall impression of the user interface?
2. Was did you think of the organization of information on the screen?
3. Was there anything you particularly liked/disliked?
4. How would you rate the complexity of the interface?

Information

- Is provided content clear, effective, fits goals (inform about past actions, current status and biases in general?),

5. Did have any difficulties to find the information needed for the tasks? If so, elaborate.
6. Was the content of the dashboard clear and understandable?
7. Were there any sections of the system that you found confusing?

Scenarios

- What are points of frustration or what works well? Which problems arose? Was it recoverable or was intervention needed?

8. Which tasks were easy to complete, and why?
9. Which tasks were difficult to complete, and why?
10. Did anything about the system frustrate you when you were using it?
11. (depending on task outcome) When a mistake happened, were you able to recover from it?

Overall

- General Feedback and points for improvement

12. Which improvements would you suggest to make the dashboard more user-friendly?
13. Does the system have all the functions and capabilities you would expect it to have?
14. Do you think that such a system would be useful in project management scenarios?
15. Are there any additional remarks, you would like to make?

FIGURE 21: Usability Interview Questions

(school/university or professional contexts). The target sample size for the usability tests was 3-5 participants, in order to identify the most prominent usability issues. The recruitment of participants took place through university-related social channels, such as whatsapp-groups, as well as word-of-mouth recruitment. During the user tests, the conversation and the performed actions were recorded using audio and screen-recording software. The used consent forms and information letters can be found in the Appendix, Figures 37 and 36.

The system designs were implemented for the usability tests as a low-fidelity prototype using figma[1]. This allowed for simple interactions with the design, such as hovering and clicking, therefore making it suitable for testing the design and the intended interactions. The used prototype for the two screens (dashboard and DSS) can be found under this link[2], or in the Appendix, Figures 34 and 35.

5.2 Feedback and Design Revisions

For the usability test, three participants with interest in, or experience with project management were recruited. This was done through through university-related social channels, such as whatsapp-groups, and word-of-mouth recruitment. The participants had either expressed interest in project management and the system, or stated that they had (limited) experience with project management from personal projects or work experience. The three participants each completed a usability test session, with an approximate duration of 30 to 45 minutes.

5.2.1 Free Exploration and Attention Patterns

During the free exploration of the dashboard, only one participant (participant B) exhibited the expected attention path according to the Gutenberg diagram. Participants A and C deviated from the expected path, with participant A noticing the average metrics at the top right first, and participant C reading the left side of the screen vertically first. An overview of the attention patterns can be seen in Figure 22.

During the exploration, the participants were asked to think aloud, in order to get some insight into which elements stand out or pull focus while using the dashboard.

[1] `https://www.figma.com`

[2] see flow 1 for dashboard, flow 2 for DSS `https://www.figma.com/proto/nkPSiUUXrQMQAnoDYVVvOs/Mockup-Thesis?node-id=0-1&t=i5yV50y65ZCoITTm-1`

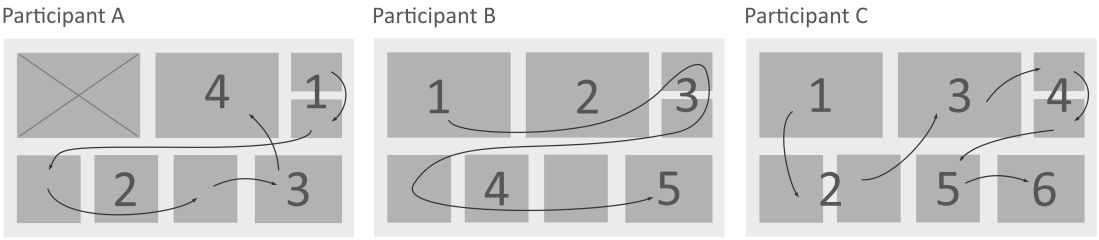

FIGURE 22: Attention Paths Usability Test

Regarding the timeline component, one participant did not mention or notice it at all during the exploration. The other participants remarked on missing captions for the axes, and mentioned some confusion about the content, showing a need for clarification of this dashboard element.

As for the detail view of project decisions, the participants mentioned liking the structured overview of project data and the option to access more detailed information (detail on demand).

For the average metrics (biased decisions per month and average financial impact), the participants remarked that they imagined those elements to be especially useful for managers and users with a financial background. Participant A commented that they were drawn to this element first when looking at the dashboard, because of the colors and bigger font-size standing out. This element was therefore slightly toned down in the design revisions, in order to not pull the user attention as much.

Looking at the three pie charts (outcome and categorization of biased decisions), the participants remarked that they think the charts would be useful, but showed some confusion regarding the categories. This showed a need for further clarification of these elements.

Regarding the information box on biases, the participants all spent some time reading the provided information carefully and comparing the provided scenario. Participant B commented that they thought this element was the most interesting part of the dashboard, as it provided the most new information to them, and that it helped them understand the detail view on the past decisions better.

5.2.2 Dashboard Tasks

Overall, all tasks except number 6 (Can you tell me the average financial impact of a biased decision?) and 8 (Can you tell me the project stage, in which most biased decisions

happen?) had a 100% completion rate amongst the three participants, making the average completion rate of all tasks 92%. The two failed tasks occurred with the same participant (B), with task 8 failing due to some miscommunication about which parts of the mock-up are clickable. This makes the overall effectiveness (completion rate) across all given tasks 100% for Participants A and C, and 77% for participant B.

The two tasks that took the longest on average were tasks 4 and 5 (Can you tell me a reason, why the last decision was flagged as biased?; Can you tell me a way to prevent risk seeking behavior?) with and average duration of 17 and 15 seconds. This was to be expected since the task required reading and reviewing more detailed text information. The tasks that took the least amount of time were tasks 6-9, since they mostly required reading a given value off a diagram or chart.

The main difficulties that occurred during the testing were: (1) confusion about the tooltip of the timeline component, (2) confusion about the order of items in the detail view (3) confusion about the difference between reconsidered and changed decisions, (4) difficulties using the drop down menu to switch between the pie charts, and (5) difficulties finding the average financial impact.

The issues 1-3 were the main points that were noted for improvement, while issue number 4 was likely due to some confusion with figma and which parts of the mock-up were functional or not. Issue 5 was likely due to exhaustion or insufficient attention during the user test, as the participant previously commented on the metric (average financial impact) during the exploration of the prototype but then had difficulties finding it during the task-based part of testing.

A detailed overview of the comments and results per task can be seen in the Appendix, Figure 39.

5.2.3 Decision Making

In order to test the decision support interface, the participants were asked to describe what they saw and noticed using think aloud (similar to the exploration of the dashboard), before making their decision using the tool.

Overall, most of the participants (A and B) read the provided data very carefully and thoroughly, following the expected top-down path. Participant C skipped more briefly over the text before making their decision. All of the participants either used the hover functions or tried clicking on links and info-buttons for more context.

The main issues that were noted during the decision making are the following: (1) the miniature charts are hard to read/understand, (2) participants are missing some context as to what the project progress is, (3) interface is missing a cue on how much reasoning/text to provide, and (4) interface should summarize options (go/kill) and their factors more clearly. All of these remarks are factors that impacted the user experience and were changed before the final implementation.

For the sake of also testing the feedback/warning screen, the system gave the warning in either case (rational/irrational choice). Of the three participants, 2 chose to reconsider their choice after receiving the feedback, while one participant confirmed their choice regardless. None of the participants noticed the pre-selected default option or the info cues about the used nudges during testing.

A more detailed summary of the comments and notes of the user test can be found in the Appendix, Figure 40.

5.2.4 Usability Interview

The usability interview was conducted following the testing of the two interfaces. Overall, the feedback was mostly positive and no major usability issues besides the previously mentioned ones were noted.

Regarding the questions on usability, the participants commented that the interface was clearly structured, informative, easy to use, and that the information load was appropriate. The participants also positively commented on the use of color, and that the sections of the dashboard were easy to differentiate. The participants also remarked that they liked the options to pull more detailed data if needed, but that the provided data overall might not be sufficient to capture such complex scenarios. The item that received the most negative feedback were the miniature graphs in the decision making screen, since they proved challenging to read and understand.

As for the questions on the provided data and information, the participants found the content to be easily understandable, with the exception of some technical terms. The main points for improvement that were remarked are the previously mentioned miniature graphs and the pie chart with the changed and reconsidered decisions. The participants remarked that this item was hard to understand and would need more explanation. Compared to that, the average metrics were noted as especially positive and easy to understand.

Regarding the questions on the given tasks, the tasks that were based on reading the charts and the information box were listed as the easiest to complete. The detail view task and the task using the reconsidered/changed pie-chart were named the most difficult to complete, in addition to the decision making task. The main points of frustration were related to the mock-up using figma and some miscommunication on which parts of the prototype work.

For the more general questions on missing features, the main feedback was that the system should provide more information on the context of the data, and that some items like the timeline and the detail view could stand out a bit more visually. The question on the perceived usefulness of the system was met with mixed results, especially since real-life scenarios are likely to be much more complex than the simple calculations presented in the screen. Participant A remarked that using the tool could be perceived as a chore if it was introduced by higher management, or that it could be met with a certain amount of resistance or spite in such scenarios. Participant B commented that it would probably be useful to see such an overview of your own decisions, but that the feedback step on the decision would probably introduce new biases, as it could be perceived as mandatory.

Overall, the usability test and the interviews showed some room for improval, especially regarding some of the charts used in the system. A more detailed summary of the interviews can be found in the Appendix, Figure 41.

5.2.5 Updated Designs

Based on the gathered feedback from the usability tests, the designs of the dashboard and the decision support screen were adapted and updated. A overview of the made changes can be seen in Table 5.1.

For the dashboard, the main improvements were (1) clarifying and highlighting the timeline, (2) clarifying the detail view, (3) toning down the average metrics to make them stand out less visually, and (4) clarifying the reconsidered/changed pie-chart.

For the decision support screen, the main points of improvement were (1) removing the miniature charts and replacing them with milestones, (2) making the reasoning prompt more detailed (how much text is expected), (3) structuring the go/kill options more clearly, and (4) making the bias feedback less leading. The updated designs can be seen in Figures 23, 24, and 25. The final implementation can be found on github[3].

[3] `https://github.com/sHustinx/DashboardWidgets`

Interface	**Change**
Dashboard	clarify timeline tooltip
	add captions to timeline
	highlight timeline visually
	clarify detail view
	clarify reconsidered/changed diagram
	tone down avg. metrics
DSS	remove miniature charts
	add milestone overview
	add expected length to reasoning prompt
	tone down bias feedback
	add more structure to go/kill factors

Table 5.1: Usability Improvements

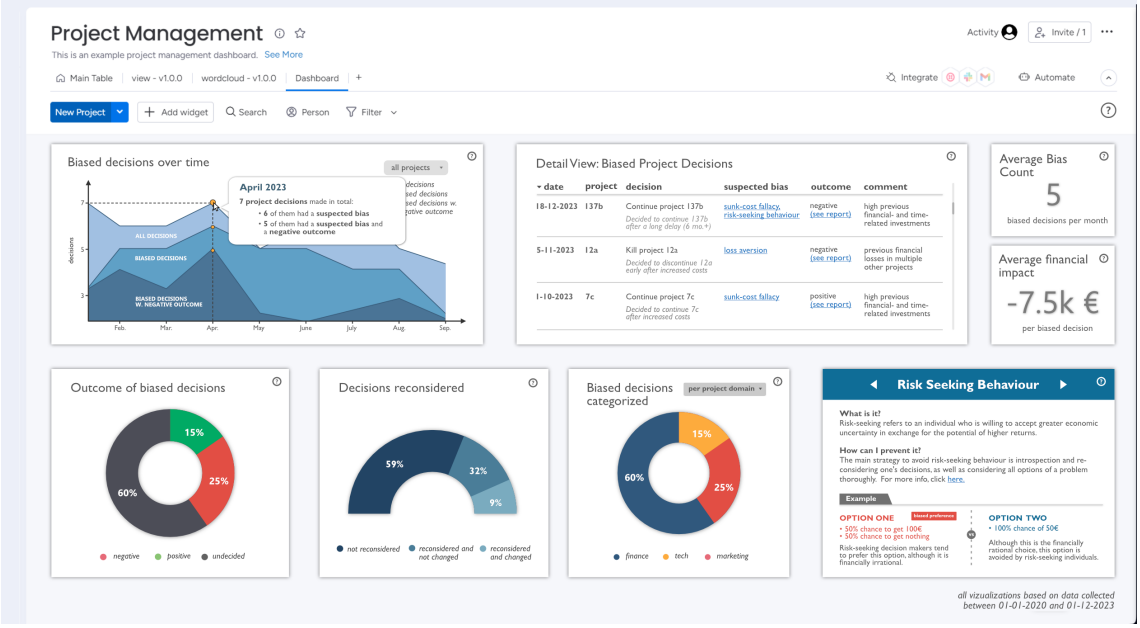

FIGURE 23: Updated Dashboard Design

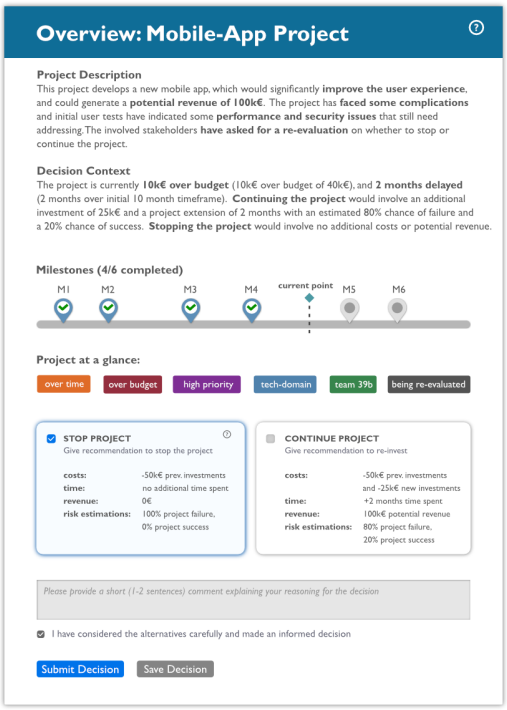

FIGURE 24: Updated DSS Design

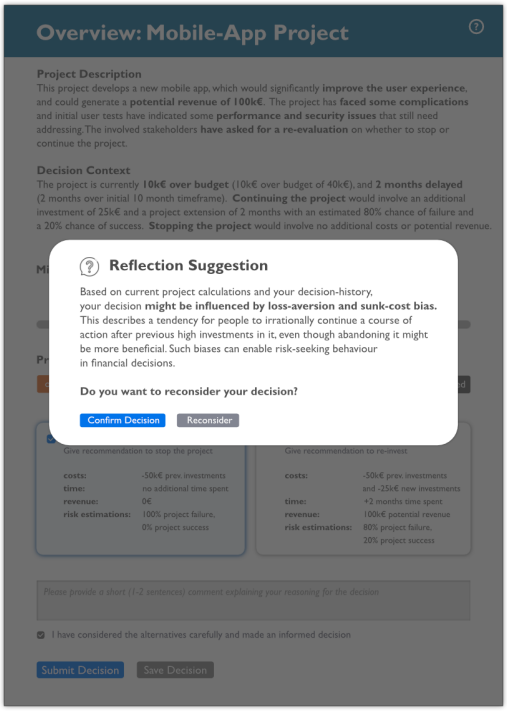

FIGURE 25: Updated Feedback Screen

Chapter 6

User Testing and Validation

This chapter will cover the objectives of the user tests, including the formulated hypothesis and research assumptions. Following this, the developed decision scenario for the user tests will be presented, along with the underlying project calculations. Finally, the conducted user tests and their results will be presented.

6.1 Objectives

To test the effectiveness of the developed system, the dashboard and the decision support screen were tested using a mixed-methods approach. To collect in-depth qualitative data about the usability and usefulness of the system as a whole, in-person user tests were conducted. To assess the user trust and effectiveness of the decision support screen through qualitative and quantitative data, an online questionnaire was used. The combination of these two research parts aims to create a balanced overview of user trust, opinion, and user behaviour with the system.

The two main research questions (see Section 1.2) of this thesis are: (1) *How to visualize past decisions and current project data in a project management context to help create bias awareness?* and (2) *How does the transparency of nudging in decision support tools affect the decision making process and user acceptance in project management contexts?* Based on this and the supporting literature, one hypothesis and two assumptions were formulated.

Hypothesis

a) *Transparent nudges have a higher level of user acceptance than covert nudges in project management decision support scenarios.*

Assumptions

Using an informative dashboard prior to making a decision using a decision support screen creates bias awareness and a motivation to correct biases.

Transparent nudges and bias feedback integrated into a decision support screen can be used to guide users towards rational and reflected decisions while preserving free choice.

Hypothesis a) is based on the findings that transparent nudges can be used to lessen the scepticism, and to strengthen the decision makers autonomy and trust in a system, especially if they are combined with benevolent nudges (mainly benefiting the decision maker) [10, 45]. This leads to the hypothesis that the transparent nudges in this system are perceived with an increase in trust and user acceptance compared to a covert design.

The first assumption is based on the literature findings showing that being aware of biases is the first step for successful debiasing, and that learning about biases (how they occur, how to avoid them, and what impact they have) is an effective strategy for debiasing decision makers [40, 14]. It is therefore assumed that this learning effect and bias awareness can be created through the use of the developed dashboard, detailing the impact, outcome, and development of biased decisions over time.

The second assumption is based on the literature findings on nudging and libertarian paternalism, namely that the applied nudging strategies can be used to effectively steer users towards rational choice and reflection [10, 57, 59]. This leads to the assumption that the applied strategies can be used to improve reflection and rationality in the context of project management decisions, without restricting the provided options.

In order to explore these assumptions and test the hypothesis, a decision scenario was developed. Due to a lack of available industry data, fictional data similar to the decision scenarios seen in behavioural economics experiments was used. In both research parts, the following scenario was used for consistency, with the options to either re-invest in the project or to stop it. The underlying project calculations can be seen in Table 6.1, and Formulas 6.2 and 6.4.

Decision Context and Scenario

This project develops a new mobile app, which would significantly improve the user experience, and could generate a potential revenue of 100k€. The project has faced some complications and initial user tests have indicated some performance and security issues that still need addressing. The involved stakeholders have asked for a re-evaluation on whether to stop or continue the project.

The project is currently 10k€ over budget (10k€ over budget of 40k€), and 2 months delayed (2 months over initial 10 month timeframe). Continuing the project would involve an additional investment of 25k€ and a project extension of 2 months with an estimated 80% chance of failure and a 20% chance of success. Stopping the project would involve no additional costs or potential revenue.

Decision Option	**Total Cost (€)**	**Expected Profit (€)**	**Expected Value (€)**	**Probability of Success**	**Probability of Failure**
Continue Project	75,000	20,000	-55,000	20%	80%
Stop Project	50,000	0	-50,000	0%	100%

TABLE 6.1: Expected Value and Project Calculations

In this scenario, the mathematically rational option would be to stop the project and minimize the losses, due to the high risk estimations (see Table 6.1, expected value). However, risk-seeking individuals might be influenced by sunk-cost bias (high previous investments) and loss aversion (not wanting the previously invested money to be lost, with no chance of profit), leading to risk-seeking behaviour and the preference to continue the project.

$$ExpectedRevenue_{Continue} = (0.2 * 100k) + (0.8 * 0) = 20k \quad (6.1)$$

$$NetExpectedValue_{Continue} = 20k(ExpectedRevenue) - 75k(TotalCost) = -55k \quad (6.2)$$

$$ExpectedRevenue_{Stop} = 1 * 0 = 0 \quad (6.3)$$

$$NetExpectedValue_{Stop} = 0(ExpectedRevenue) - 50k(TotalCost) = -50k \quad (6.4)$$

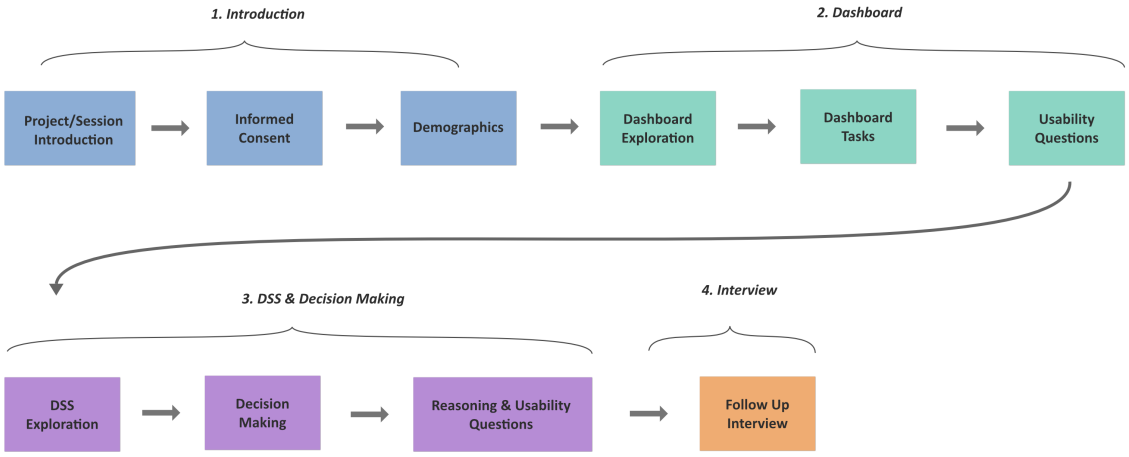

FIGURE 26: Structure In-person User Tests

The target group for the in-person testing and the online questionnaire are participants who either have experience with, or interest in project management. The participants must be over the age of 18, and able to consent to this type of research. The recruitment of participants was done through university related channels, such as whatsapp groups, canvas, and e-mail. In order to reach an audience that had some level of experience with project management, a class of project management students were approached specifically, but it was made clear that any participation would be voluntary and anonymous.

6.2 In-person User Testing

This section will cover the in-person user tests. First, the used methods and the structure of the research will be outlined. Then, the gathered results will be presented, along with any potential limitations of the research.

6.2.1 Methods

The objective of the user test was exploring how participants interact with the system and whether or not the used nudges and the provided information might prompt bias awareness or debiasing. That made the two assumptions on bias awareness and transparent nudging the main focus for the in-person user testing.

The user tests focused on gathering in-depth qualitative data. This was done through the use of semi-structured interviews, and think aloud data collected during the use of the system. Since the user tests were estimated to take quite long (roughly 45-60 minutes) and

to produce a lot of qualitative feedback, the target sample size for this research step was set to 5 participants, in order to allow for more in-depth testing and evaluation. The recruited participants were invited to project rooms at the university, and received a drink and sweet treat for their participation.

The structure of the experiment followed a four part structure: (1) introduction and demographics, (2) dashboard usage, (3) decision support screen usage, and (4) a follow up interview on the system as a whole. A visual overview of the structure can be found in Figure 26.

First, the participants were given a brief introduction about the project and the session structure, before reading and signing the provided information letter and informed consent form (Appendix, Figures 42 and 43). The participants were informed that their sessions were audio- and screen-recorded for further analysis and transcription. Then, some demographic data on the age and experience of the participants with project management (what level of experience and in which contexts) was collected. Finally, the participants were introduced to the system (dashboard and DSS implemented within monday.com) and provided with the previously described decision context (Appendix, Figure 44). During this, the emphasis was on the decision context and the structure of the tool, without mentioning anything related to biases. This was done in order to allow the participants to give more unfiltered reactions and feedback, and to avoid the introduction of new biases into the process.

For the second part of the user test, the participants were first asked to explore the dashboard on their own using think aloud. This was meant to collect qualitative data on the path of attention and to allow the participants to familiarize themselves with the dashboard and its functionalities at their own pace. After that, the participants were asked to complete three tasks based on the project requirements (Section 3.1), in order to see how they would use the system to find data about the development of biased decisions over time, the impact and outcome of those decisions, and detailed decision data. Here, it was recorded which elements of the dashboard the participants used to complete the tasks, and whether or not they were able to find the relevant data. The tasks were designed to allow for some flexibility as to which components can be used, so that the participants could use different dashboard components to inform themselves. For example, for the first task, participants could use the timeline component or the detail view to access the relevant data.

Dashboard Tasks:

1. How would you use the dashboard to inform yourself about the development of your decision history over time?
2. How would you use the dashboard to inform yourself about the average outcome and impact of decisions?
3. How would you use the dashboard to inform yourself about the last made decision and the details of that decision?

After completing these tasks, the participants were asked to answer some open-ended questions about the usability and usefulness of the dashboard, since those are the main two factors impacting dashboard acceptance [30]. The questions for this were partially adapted from the UMUX (Usability Metric for User Experience) [22] questionnaire, which is a lightweight alternative to the SUS questionnaire and aims at assessing the general usability of a system.

Dashboard Usability:

1. What was your general impression of the dashboard?
2. Did you find the dashboard data useful and adequate for completing your tasks, and why or why not?
3. Were there any points of frustration or difficulties you encountered?
4. Did you think the dashboard was easy to use or not?

For the third part of the user test, the participants were asked to explore the decision support screen using think aloud, before making their choice using the system. This part of the experiment gathered more qualitative data about the user perception and thought process during decision making. It also collected the decision data itself, so which option the participants chose, whether they reconsidered their choice, and which interface elements were used to make the decision. Following the decision, the participants were asked a number of follow-up questions related to their reasoning and the perceived usability.

DSS Questions:

1. Why did you choose this option, and which elements influenced you?
2. (if they picked the irrational option and received feedback) Did you reconsider your choice after receiving feedback and why or why not?
3. Did you find the decision support screen helpful for making your decision, or not?
4. Were there any specific elements of the support screen that stood out, either positively or negatively?

The last part of the user testing consisted of a follow-up interview on the system as a whole, so the dashboard and DSS in combination. This part's main purpose was to assess whether using the dashboard did influence the decision making or not, as well as gaining an overview of the perceived usefulness. After the interview, the participants had the chance to ask any remaining questions or give final remarks, before concluding the experiment. A detailed overview of the used research script, including the asked questions and time estimations can be found in the Appendix Figure 45.

Follow-up Interview Questions:

1. How would you describe your experience with the system as a whole?
2. How would you rate the usefulness of the system for project managers?
3. Are there any challenges or drawbacks that you would expect with using this tool?
4. Were there any features or sections of the system that you found confusing or unnecessary?
5. Are there any additional features/functionalities that would improve the system in your opinion?
6. Did using the dashboard influence your decision making process? If so, how?
7. Do you think it is important for project managers to be aware of their own biases when making decisions? And why or why not?
8. Was there any information you learned during the use of the system, that was surprising or new to you?

The gathered data from the interviews was transcribed and evaluated through a deductive thematic analysis, following the guidelines by Braun and Clarke [7, 41]. This thematic analysis was then used in combination with the collected decision data and the remarks and observations made during the testing, in order to gain a more detailed overview of the effectiveness of the system in regard to the research questions. This part of the research was mostly exploratory in nature, aiming at understanding how the participants interact with the system and whether they are influenced by certain elements of it.

6.2.2 Results

The in-person user tests were done with five participants (referred to as participants A to E) in total, with ages ranging from 24 to 29. One participant listed their experience level as beginner (<1 year experience) and the remaining four participants categorized themselves as intermediate (1-3 years of experience), with two of the intermediate participants having roughly one year of corporate project management experience. The rest of the participants stated that they mainly use project management for time- and resource planning of more elaborate university-, and personal projects.

Two of the five participants (A and C) chose to continue the project, receiving the recommendation to reconsider their choice, which both of them rejected. The remaining three participants chose to stop the project without any feedback, exhibiting risk-averse behaviour. An overview of the choices and reasoning per participant can be seen in Table 6.2.

During the free exploration of the dashboard, three of the five participants (A, B and E) showed the intended pattern of attention, following a ‚Z‘-pattern across the screen (see Figure 27). Participants C and D deviated from the expected pattern by reading the left half of the dashboard in a top-down manner before moving on to the other dashboard components. This showed some improvement compared to the preliminary usability tests.

Overall, the interview data showed some reoccurring themes that fall under three main categories of (1) interaction and content, (2) general system usage, and (3) biases and nudging (see Figure 28).

Choice	Participant	Reconsidered	Reasoning
Continue	A (intermediate, age 29)	no	*The project is almost done, marked with high priority, and high revenue if it gets completed.*
	C (intermediate, age 24)	no	*I trust our team.*
Stop	B (intermediate, age 24)	-	*I think the risks are too high, and even though we are far into the project, it does not seem to have a good outcome.*
	D (beginner, age 27)	-	*Almost twice the budget in combination with an 80% chance of failure leads me to believe stopping the project is the best idea.*
	E (intermediate, age 27)	-	*Already 50k spent. Continuing is an extra 25k, but with a 80% chance of failure, so probably this means just spending 25k extra without earning anything back.*

TABLE 6.2: Overview of Choices and Reasoning User Testing

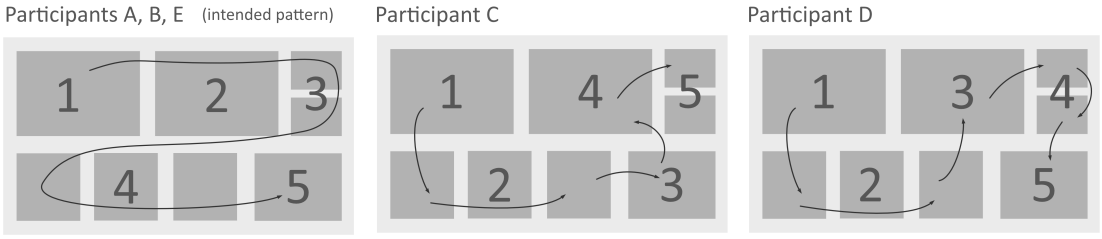

FIGURE 27: Attention Patterns User Tests

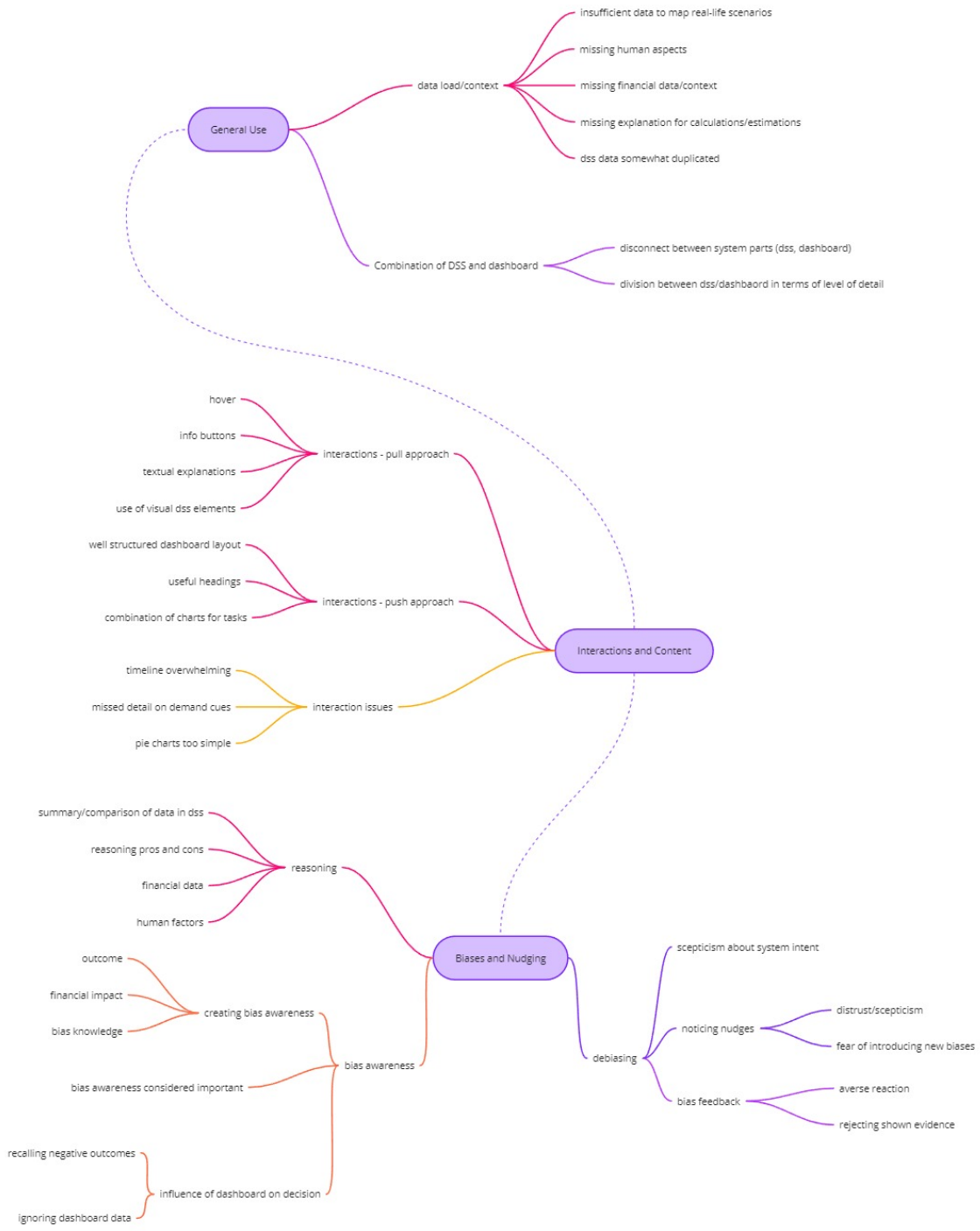

FIGURE 28: User Test Themes

Interaction and Content

As for the provided content and the interface interactions, as well as the overall usability of the system, the feedback was mostly positive. Frequently mentioned aspects were the structured layout of the dashboard, and the descriptive headings of each section and chart, which helped the participants understand the content better. The participants also showed no problems completing the given tasks, with most of them combining data from multiple charts and comparing the data between charts to give informed answers.

> B: *"I think it's very nicely structured, and I like that everything is in these rectangles. So everything looks separated from each other."*

> D: *"I think it looks good. It looks clean. I know where to find everything. All the titles of the segments are pretty self explanatory, right? So with the tasks that you just gave me, I instantly know where to look. "*

The participants also had no trouble finding and using most of the provided interactive elements, such as hovering for more data or using the information buttons. The feedback on these interactions, especially on the information buttons was positive, as can be seen in the quote below.

> A: *"I liked the information, for example [...] for outcome of biased decisions I was hovering over this question mark to have a more detailed explanation. And in general, I like that. [...] It doesn't feel like you're left with the dashboard, you can still kind of ask questions or look back at something."*

The main interaction issues that occurred during testing and that were remarked on during the interviews, were with the timeline component, the pie charts, and some detail on demand functions. The component that caused the most confusion was the timeline of biased decisions, with the participants struggling to understand the presented data and not realizing it was possible to interact with the graph (hover and filter) to access more detailed data. The same occurred with some of the participants struggling with (or not noticing the option of) opening a more detailed overview of each project in the detail view. Finally, the last frequently occurring theme during the discussion of the dashboard, was the bottom row of pie-chart diagrams. Participants remarked on a lack of complexity in these charts and that they would be more useful when combined with other metrics.

A (about points of frustration): *"With the very first diagram on the top left (timeline). So for biased decisions over time, because in the beginning, I did not entirely get how it really works. Because in the beginning I didn't see that you can actually hover over it to have more information directly on the diagram, which is nice, I think. [...]"*

D (about reconsidered pie-chart): *"Because on its own I think this information is not so valuable. But if you link that to the result of the reconsiderations, that might be more interesting."*

E (about pie-charts): *"Maybe if these (pie charts and detail view) were combined [they would be more useful]. And if you would have like a financial chart, then you wouldn't need those either. "*

General System Use

The main themes that were found within the category of system use were the data load and context, and the combination of the dashboard and the decision support screen.

While participants commented that they liked how the two parts of the tool complemented each other in terms of scope (detailed summary in the DSS and executive overview in dashboard), almost all participants remarked that they felt a lack of connection between the two.

C: *"The dashboard was really nice because it tells me the whole picture. And this one (DSS) tells me like the actual groundwork. I think if there was some more connection between the dashboard and this part (DSS), then it would help a lot in really stating clearly that you should not be biased. [...] Because when I was doing the second system (DSS) I had already forgotten about the first system (dashboard)."*

E: *"I think to make the decision I just made, I don't really need this (the dashboard), but if you would do more [decisions] I would like an overall view for the projects, then I would choose this (the dashboard)."*

As for the data load and the context of the provided data, the most common remark was that the provided data would not be sufficient to realistically match a real-life scenario.

Related to this, some participants were missing more detailed financial data in the project overview of the dashboard, and more information and transparency on how the estimations and calculations are created by the system. Using the decision support screen, participants appreciated the summarized overview of the most relevant data, but also remarked that the data was somewhat duplicate (provided in the textual description and visual overview of the choices). The last theme that came up related to the data and context was the lack of human impact shown in the decision support screen: participants remarked that they were missing some information on the actual people that their decision affects, like the related project team.

> B: *"At the project at a glance, I saw the team that was working on it and I thought if I would click on it, maybe it would show me the team. Although I think that would have probably also made me rethink that decision, because if I would have then seen names of people who are maybe then unemployed after that it would have been hard."*

> D: *"So I think we were already two months overtime and then two months extra, but I feel like that's difficult to say in this scenario where it's not like an actual company. That misses a lot of context, I would say. "*

> D: *"I guess something that would be interesting is to get like more context behind these risk estimations. Like what do you base these 80% project failure on?"*

> E: *"Do we see the full costs of the projects (in detail view)? Because 50K over like 10K is a lot, but if it's like a project of 2 million, then it's not."*

Biases and Nudging

The themes categorized under biases and nudging mainly center around bias awareness, the observed reasoning, and the effectiveness of the debiasing.

Overall, the two participants that chose to continue with the project (A and C) were the only ones to actually notice the transparency cues about nudging, or the nudges (pre-selected default) themselves. Participant D did notice the cue about the intent of the system (improving rational choice), but only noticed the default nudge during the follow-up interview. Similarly, all five participants said that they consider bias awareness to be very important in the context of decision making, but only the three participants that made

the rational choice (B, D, E) said that the use of the dashboard influenced their decision making process in some form. This shows some disconnect between the opinion on bias awareness and the actual willingness to accept feedback and reconsider the choices made. Regardless of that, all five participants showed some degree of bias awareness during the dashboard use, either by commenting on the high impact and negative outcome of different biases, or learning about and commenting on specific biases using the dashboard. This indicates that the use of the dashboard previous to making the decision did support bias awareness for the participants, but did not necessarily motivate all of them to correct those biases. A collection of such remarks and comments indicating bias awareness can be seen below.

> B: *"I can also see the financial impact (of biased decisions), which is very big."*
>
> C (about biased decisions in detail view): *"Seems we're pretty stubborn, like we often don't want to take our loss. And also seems pretty naïve. And all the projects are going pretty horrible from the look of it."*
>
> D: *"Like this, the suspected bias types, which is why it's nice that it's optional to get more information about it. [...] This seems relevant for my future decision. For example this one. (looks at past decision and bias in detail)"*
>
> D: *"Average financial impacts minus 7.5k. Okay, this suggest that like bias is always negative, right?"*
>
> C: *"I would look at the outcome of biased decisions by chart, and then it tells me that yeah, we should not make the biased decisions because they have bad outcomes."*
>
> D: *"Because now I've made this decision to stop the project. I guess in that case there would be no bias involved. Maybe at least from the list that I saw before, it didn't seem like any fit to making this decision to stop the project. But I could think of some that would be related to continuing the project."*

The participants that said that using the dashboard did influence their decision making process in some way (B, D, E) mostly mentioned recalling the negative effects of biases shown in the dashboard as factors in their process, and remarked on the importance of bias awareness in the follow-up questions. Some supporting quotes of this can be found below.

B: *"I know that sometimes when I was reading the word decision, I was thinking about like the biased decisions and know how high they (the losses) were at some point. So I was like, okay, maybe that's good to take into account."*

D: *"I mean objectively, being aware of those biases leads to you getting affected by them less. Therefore, likely making better decisions, right? [...] This sunk cost fallacy, for example, exists for a reason and therefore probably if you're aware of it, you can earlier recognize that you are affected by that and therefore make a better decision. Even if you continue doing it, at least you are aware of it, right?"*

E: *"Yeah, a little bit. (on whether dashboard influenced decision) Because I was already thinking that there is so many negative outcomes. I was already assuming that the project was gonna be probably negative. "*

For the two participants that chose to continue the project, although they showed some level of bias or self-awareness (A saying they picked the *risky* choice, and C saying they know they are biased), the transparent nudges and the bias feedback seemed to have adverse effects instead of improving rationality. For participant A, the bias feedback and the pre-selected default led a breakdown of communication, where the participant read the feedback but did not make the connection to their own decision, instead confirming their choice more confidently. Only during the follow-up interview (during the question on the importance of bias awareness), did the participant have a change of mind and said that they perhaps should have reconsidered their choices.

Participant C had already remarked during the use of the dashboard, that they felt like the provided information on biases was confrontational and paternalistic (see quotes below).

C (about the information box on biases): *"To me it felt a little bit... Condescending is not the right word, but it's trying to teach you too much. I felt like it overstepped its bounds. "*

C (about average metrics): *"OK, they even give us the amount of biased decisions we do on average, which is pretty high. Yeah, It's a bit confrontational"*

During the decision making, the transparency cues and the pre-selected default then led to an increased distrust of the system and its intents. When receiving the bias feedback, the

participant said that they were aware that their decision was biased (showing some degree of bias awareness), but not willing to reconsider their choice (showing a lack of debiasing motivation). This led to the decision being confirmed with increased conviction. Therefore, the applied nudges and the bias feedback did not lead to a reflection or reconsideration of biased choices. In addition to that, the transparent nudges were met with rather high levels of scepticism and distrust during the user tests, which negatively impacts hypothesis a) on the user acceptance of transparent nudges.

A comparison of the reactions and comments made by the two participants in regard to the bias feedback and the importance of bias awareness can be seen in the quotes below. These comments also show the disconnect between the opinion on debiasing and bias awareness in general, compared to the actual reaction when faced with bias feedback.

Participant A

A: *"Okay, this (the bias feedback) is interesting because this suggestion is not related to my decision, I think. It's saying that I made a decision and that it might be influenced by loss aversion, but I made the risky choice.. So I confirm, because we don't get influenced by anchors."*

A: *"I did not reconsider because the advice that was made was not supportive of the choice that I made anyways. [...] So I was kind of against that advice from entering in it and that's why I chose to not reconsider the decision."*

A: *"And it (the stop option) was already pre-selected so, the pre-selection might give a bias to people who do not pay attention to detail. Because it's kind of nudging the decision, which might be a bit dangerous."*

A: *"No, it is important (being aware of your own biases), especially because it's known how people get biased, and it's important to create awareness. Also, maybe like in this tool... Therefore, indeed, maybe I should also reconsider it."*

Participant C

C: *"This tool shows me I'm probably biased and I already know that. But then I'm also like thinking of the actual people involved, which to me is more important than being biased or not. So, I confirm the decision."*

C: *"It's mainly because I'm a bit stubborn and also because to me, the system itself was very much about numbers, not about people. And then for it to already recommended an option without knowing the actual people, it made me feel even more stubborn to actually change it."*

C: *"It's because it felt like it (the system) was telling me to do my job better, and that didn't feel nice. So, I got more stubborn. [...] It did help me with like weighing the pros and cons. But then it also influenced me a lot by like saying what's right and what's wrong."*

C: *"The concept of having a tool that says avoid being too biased or what bias does to your financial gains and such, that's very useful. [...] I think for society as a whole it's important that everyone really thinks about such biases."*

As for the reasoning, the participants that chose to stop the project (B, D, E), all listed the sub-optimal current status of the project, the high risk estimations, and the total costs and revenue as their main points. The two participants that chose to continue the project (A, C) mainly empathized the sure loss of the previous investments in case of stopping, compared to the (small) chance of high returns in case of success. These participants therefore showed classic signs of loss aversion, sunk cost bias, and risk-seeking behaviour in general, which can also be seen in their reasoning and comments on why they chose this option. One

participant (C) also mentioned taking the human aspects (the project team) into account when making the decision, and showed increased suspicion towards the credibility of the decision support system.

> A: *"I would continue with the project because we don't give up. [...] Mostly, the potential revenue influenced it [...] Because it's a lot spent, yes, but if we stop the project then the cost and money is definitely lost. Now we have to spend more, yes, but the revenue also has a promising plus of 25K. There is a high risk for failure, but there's still some chance for success."*
>
> B: *"I chose to stop the project and because it, especially in the decision context, it said that we're already over budget. We already have a delay and the chance of failure is 80% and there is only a chance of 20% of having success with this project. So I thought that that (stopping) would be the better decision. Maybe we could use the money to focus on a different project that can actually succeed."*
>
> C: *"I think it's strange that the tool is able to estimate the risk even though they don't know the people. So I'm a bit skeptical about that already, that they recommended to stop it.*
> *Yeah, with this option I can make more money. With this option I will only lose money. (continue vs stop) Yeah, I would want to continue the project [...] I think I trust our team."*

6.2.3 Threats to Validity

Possible threats to the validity of the user tests mainly involve the target group and sample size. Since this project has no industry partner that could supply data or requirements, the system was tested using fictional data, and evaluated using a relatively small group of participants that had limited or moderate experience with corporate project management. This could impact the feedback on the effectiveness and usefulness of the system, as well as the results and use of the system in general.

6.3 Online Questionnaire

This section will discuss the conducted online questionnaire. First, the used methods and the structure of the questionnaire will be outlined. Then, the gathered results will be presented, along with any potential limitations of the research.

6.3.1 Methods

Objectives

The main objective of the online questionnaire was testing the decision making screen with a larger group of participants (compared to the 5 participants for in-person testing), and to gain insights into the user acceptance of the used nudges in the decision architecture.

The questionnaire was therefore mainly used to test hypothesis a) on whether or not transparent nudges are perceived more positively than covert ones, by comparing the user trust of two different designs (transparent and covert nudges). In addition to that, the questionnaire also gathers information on whether or not the participants feel prompted to reflect on their choices before finalizing their decision.

The target sample size for this research was 20 or more participants, in order to gain a broader overview of opinions, and to extend the data collected during the in-person tests. The questionnaire took an estimated 5 to 10 minutes and could be completed online at the participants' convenience. The questionnaire could be completed on mobile or desktop devices, but using a desktop device was recommended during recruitment, as the questionnaire contains detailed images. The participation in this research was voluntary and anonymous, as no e-mail addresses or other identifying data were collected. The provided information for the informed consent can be seen in the Appendix, Figure 46.

There does not seem to be a standard approach for measuring the user trust and acceptance of nudges specifically. Similar projects that assess the user trust of nudges either used custom scales and statements specific to the use case, or a combination of general user trust scales (like the TAM) with more specific custom questions [37, 18]. Some promising approaches, like the nudge acceptance model [27] or the DINU model (digital nudging process model) [46], are still purely theoretical, without any specified scales or questionnaires. Therefore, this research design uses a similar approach to [37, 18], combining existing scales of user trust and acceptance with more specific questions on nudging.

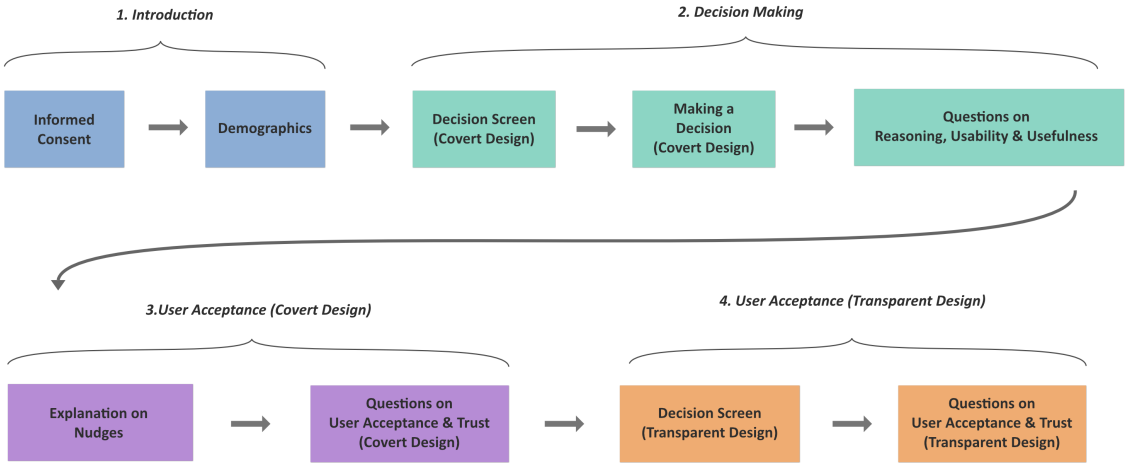

FIGURE 29: Structure Online Questionnaire

Structure

The questionnaire had three main parts: (1) making a decision using the decision support screen with covert nudges, (2) rating the trustworthiness of the design with covert nudges, and (3) rating the trustworthiness of the design with transparent nudges, comparing it to the covert design. The research used a within-subject design, in order to allow for a direct comparison of the two designs, and to accommodate a smaller sample size. A detailed overview of the questionnaire structure and the time estimations can be found in the Appendix, Figure 52. A visual overview of the questionnaire structure can be found in Figure 29.

First, the participants were asked to complete a short demographic section with questions about their age and level of experience with project management (Appendix, Figure 47). Then they were provided with the decision context and an image of the decision screen with covert nudges, and asked to make their decision (Appendix, Figure 48). This decision was then followed up by a survey section with questions about the reasoning behind the choice and the perceived usability and usefulness of the screen (Appendix, Figure 49). For this, the statements 1 and 2 were adapted from the TAM (Technology Acceptance Model) [17], and balanced with a more negatively worded statement (3), and two additional statements on the quality of the provided information. The responses were collected using a 7-point Likert scale (strongly agree to strongly disagree), and an additional open text field was added for further feedback.

Demographics Section:

1. What is your age? (with provided age ranges)
2. How would you rate your level of experience with project management? (Little/no experience; Beginner; Intermediate; Professional)
3. In what contexts do you use project management? (School/University; Work; Other)
4. Can you briefly elaborate on your experience with project management? (Open text field, not required)

Decision Section:

1. Which option would you choose? (Stop; Continue)
2. Why did you select this option? (Open text field)
3. Are there any elements or factors that particularly influenced your decision? (Open text field)

Usability and Usefulness Statements:

1. I think that using such a system would make it easier to make go/kill decisions
2. I would find this system useful when making go/kill decisions
3. I think this system would complicate making go/kill decisions
4. I think the provided information is relevant for making go/kill decisions
5. I think the provided information is insufficient for making go/kill decisions

Following the decision section, the participants were provided with a short explanation on nudging and an image highlighting the used nudges in the system (Figure 30). They were then asked to rate the previously used system (DSS with covert nudges) in terms of user acceptance and trust (Appendix, Figure 50). For this, the statements 1 to 3 were adapted from the Scale of Trust in Automated Systems [31], which is one of the most commonly used measurements for human-automation trust [9]. In addition to these statements, two additional ones regarding the perception of the systems benevolence (who do the nudges benefit?) and the user acceptance of nudges in the specific scenario were added. To create

a consistent scale across these five statements, a 7-point Likert scale (strongly agree to strongly disagree) was used, as recommended for the Scale of Trust in Automated Systems. In addition to these statements, a field for open-ended feedback was provided, in case the participants had more specific remarks about the user acceptance of the system.

User Trust and Acceptance Statements:

1. I think this system is deceptive or manipulative
2. I am suspicious of this system's intents, actions, or outputs
3. I can trust this system
4. I think this system has my best interest in mind
5. I think the applied strategies (nudges) are acceptable in this specific use case

After rating the covert nudges, an alternative transparent design with information cues and explanations was shown (Figure 31). The participants were then asked to compare and rate the trust and acceptance of this transparent design using the same questions from the previous section (Appendix, Figure 51).

Following the data collection, the questionnaire responses were analyzed and visualized with python, using numpy[1], SciPy[2], matplotlib[3], and seaborn[4] as the main software for evaluation. The statistical analysis of the results can be found in Section 6.3.2.

6.3.2 Results

Participant Demographic

The online questionnaire was promoted over a duration of two weeks, and was completed by 25 respondents. Of these 25 participants, 56% were between 25 and 34 years old, and the remaining 44% were between the ages of 18 and 24.

As for their experience with project management, most participants (48%) categorized themselves as intermediate (1-3 years of experience), followed by 24% categorizing themselves as beginners (<1 year of experience) and 16% as professionals (3+ years of experience).

[1] `https://numpy.org`
[2] `https://scipy.org`
[3] `https://matplotlib.org`
[4] `https://seaborn.pydata.org`

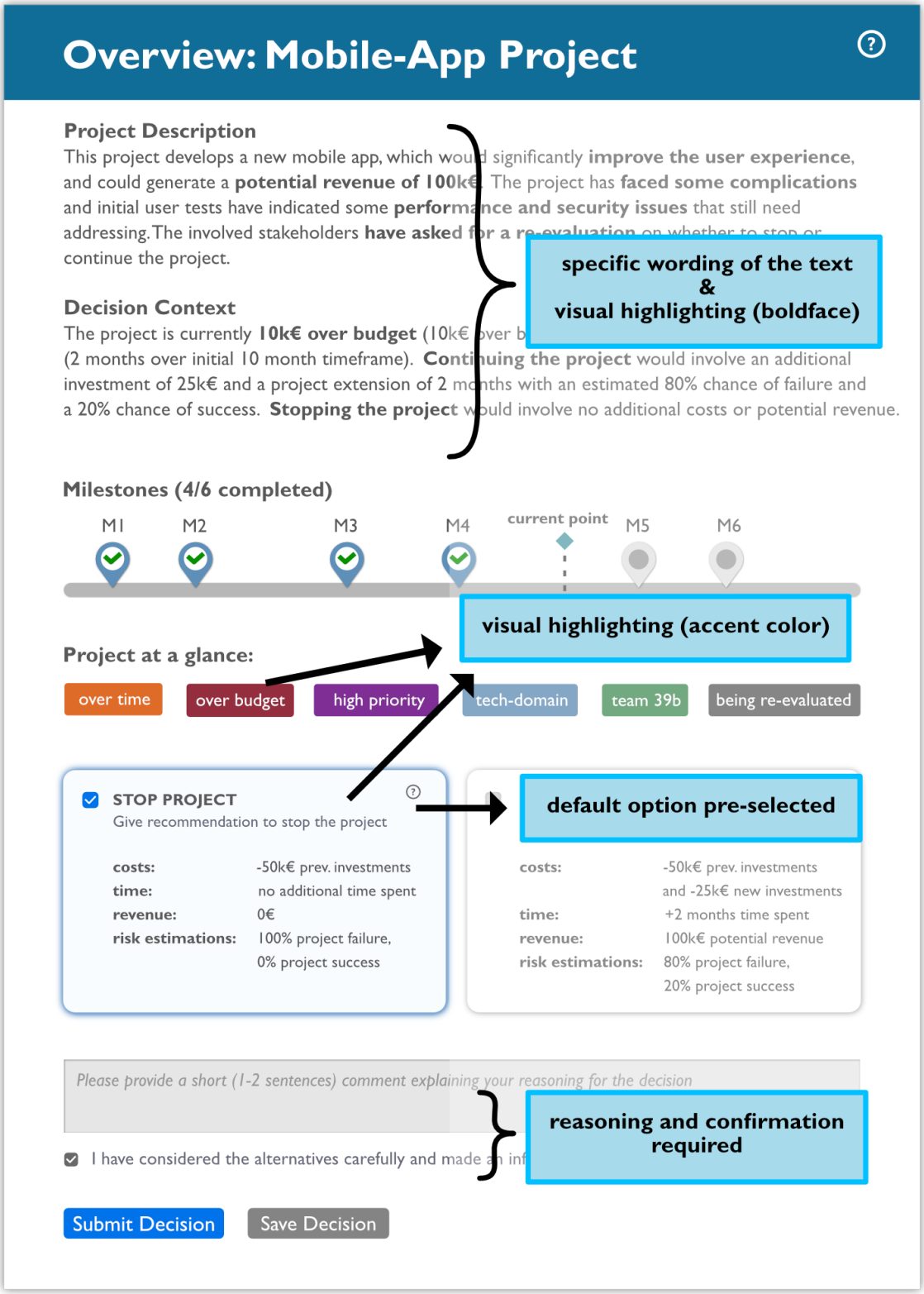

FIGURE 30: DSS with Highlighted Nudges

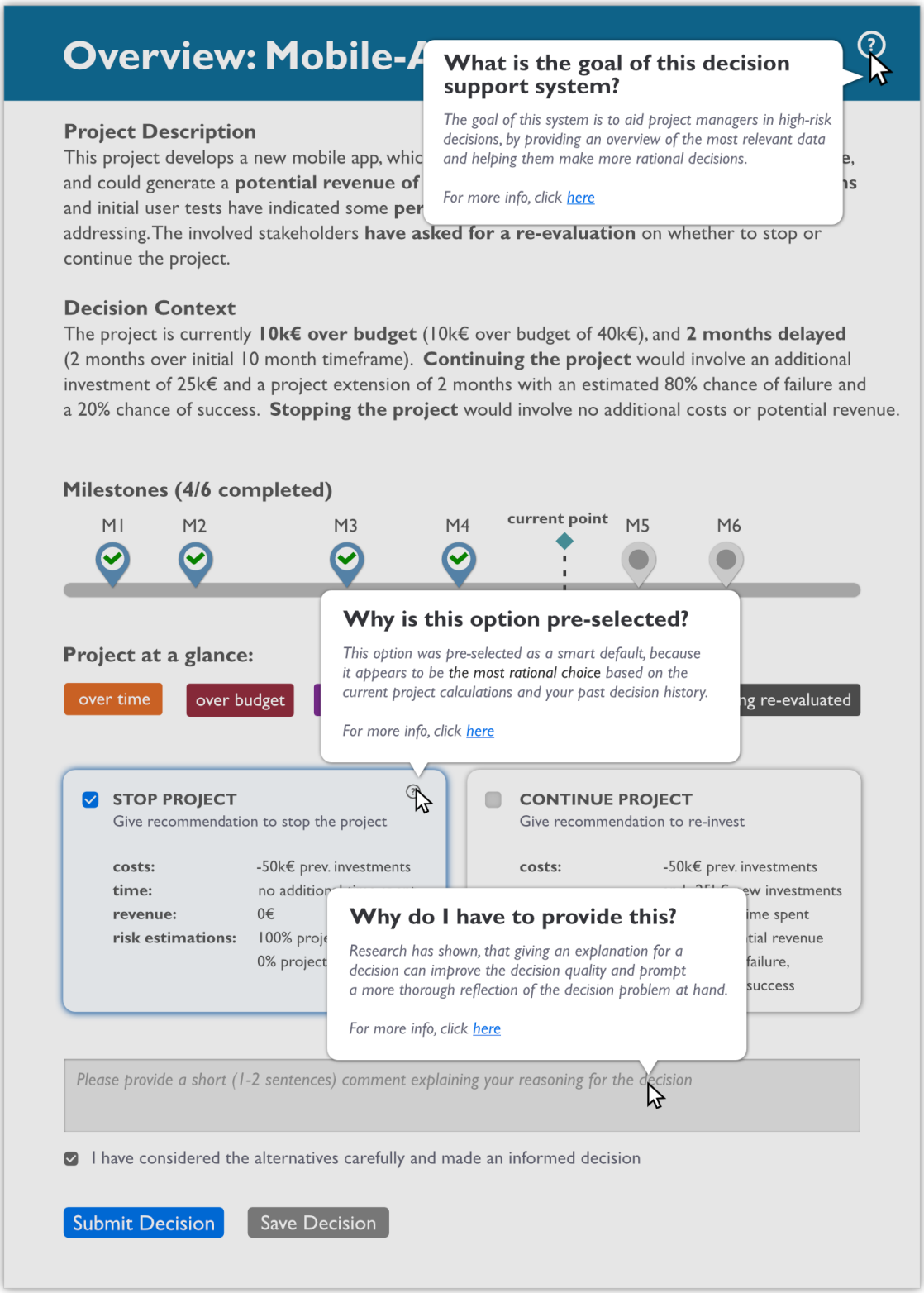

FIGURE 31: DSS with Info Cues

The remaining 12% responded with having little or no experience with project management (see Table 6.3).

The majority of participants (88%) use project management in university and school related contexts, and 48% of participants stated that they use project management in professional contexts. One participant added to the category of 'other', by listing that they use project management for volunteer work and in student activism related contexts (see Table 6.4). From the participants that use project management in professional contexts, the majority state having corporate or start-up work experience, especially in project management positions related to software development, which matches the provided decision scenario. A comparison of the results for participants with low and high levels of experience can be found in Section 6.3.2.

Experience Rating	***Number of Participants (n=25)***
Little to no experience	3 (12%)
Beginner (less than 1 year experience)	6 (24%)
Intermediate (1-3 years experience)	12 (48%)
Professional (3+ years experience)	4 (16%)

TABLE 6.3: Level of Experience with Project Management

Decision Making and Usability

Of the 25 participants, 80% made the decision to stop the project, and 20% chose to continue the project. Of the 5 people that chose to continue the project, 3 had rated themselves at a intermediate level, and 2 at a beginner level. Of the 20 people that chose to stop the project, 4 were rated as professionals, 9 as intermediate, 4 as beginner, and 3 as

Experience Context (multiple answers possible)	***Number of Participants (n=25)***
University/school	22 (88%)
Professional contexts	12 (48%)
Other contexts (activism, volunteer work)	1 (4%)

TABLE 6.4: Contexts of Project Management Usage

having little or no experience. An overview of the experience level per choice can be seen in Figure 32

From the 20 participants that chose to stop the project, 16 mentioned the unfavourable risk estimations (80% of failure) as the reasoning behind their choice, with 3 of them adding their own expected value and profit calculations. In addition to that, 6 participants mentioned that they also took the surrounding factors of the project (requirements, time investments, project status) into account when making their decision. Two participants mentioned that the money could rather be spent on more promising investments with higher returns or better chances, showing awareness of risk preference and sunk costs. A quote of such a response can be seen below.

> P10: *"The odds are very unfavourable. Doubling down with 20% chance of making profit is not worth it. Common mistake done in stock markets or gambling where people want to recover their lost investment with another risky investment and loose even more."*

Of the five participants that chose to continue the project, all five mention the potential revenue as the reasoning behind their decision, with four explicitly mentioning that the already invested money would otherwise be lost. An response like that can be seen in the quote below, which shows a classic example of loss aversion and sunk cost bias. Two participants also mentioned that they took the surrounding factors of the project (late project stage, possible improvement of user experience) into account when making their choice.

> P18: *"Because two months isn't a long time and it would be bad if 50k are lost without possibility for revenue."*

As for any interface elements that particularly influenced their choice, the participants that chose to stop the project mentioned the overview of cost and revenue, and the overview or risk estimations. The overview of labels in the at-a-glance section was also mentioned positively. The participants also remarked that it was too unclear how the risk estimations were made, with one participant mentioning that they were missing more information like market research, management briefs, milestone data, and other options than go and kill. The participants that chose to continue the project often remarked on the overview of the project advancement (milestones and data on current issues) as a factor in their decision.

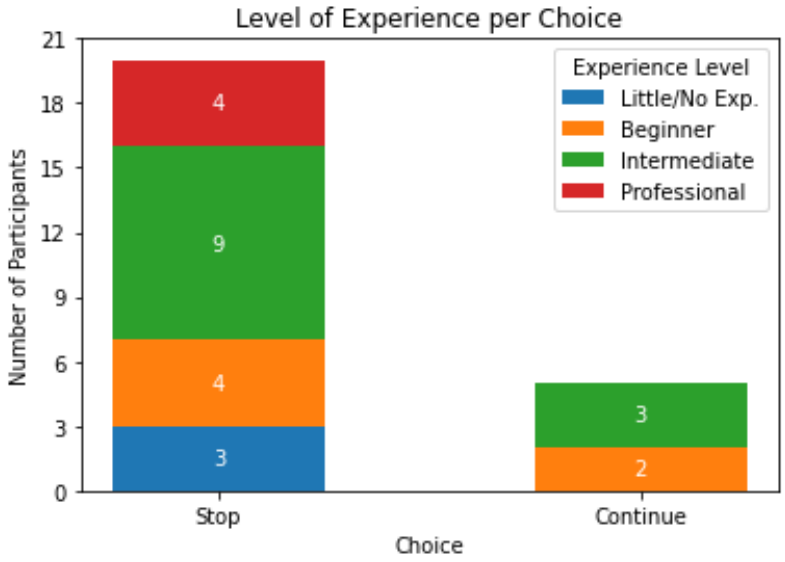

FIGURE 32: Level of Experience per Choice

Statement	**Median**	**Mode**	**Mean**	**SD**
1) I think that using such a system would make it easier to make go/kill decisions	6	6	5.2	1.47
2) I would find this system useful when making go/kill decisions	6	6	5.4	1.44
3) I think this system would complicate making go/kill decisions	3	2	2.84	1.43
4) I think the provided information is relevant for making go/kill decisions	6	6	5.36	1.46
5) I think the provided information is insufficient for making go/kill decisions	5	5	4.16	1.79

(7-pt Likert Scale, 1=Strongly disagree, 7=Strongly agree)

TABLE 6.5: Descriptive Statistics of Usability Statements

Regarding the statements on usability and usefulness, the overall feedback was positive (see Table 6.5). For statements 1 and 2 on whether the system was perceived as useful, the mode and median of the results came to a 6 (agree), and the mode for statement 3 (whether the system would complicate the process) came to a 2 (disagree). Regarding the relevancy and usefulness of the provided data, the mode for statement 4 (information is relevant) came to a 6 (agree) and for statement 5 (information is insufficient) to a 5 (somewhat agree). The last statement also had the highest standard deviation (1.79) with a large divide on opinions. Charts with a more visual overview of distributions can be found in Figure 53, Appendix.

The most frequently mentioned feedback that was given in the optional free-text field on usability was the need for more detailed information in order to match real world scenarios (mentioned in 7 of the 12 feedback responses), including team information, development tickets, other current projects, and more detailed progress data. In addition to that, the milestone and project at a glance tags were remarked on as not providing enough information, which could have been influenced by the fact that some of the designed interactions (like hovering for more data) were not available in the online questionnaire.

User Acceptance of Covert Nudges

The section of the questionnaire on user trust and acceptance of covert nudges produced more varied results, with some divide as to whether the system is seen as trustworthy or not. (See Table 6.6)

The first two statements on manipulation and suspicion resulted in medians of 4 (neutral) and 3 (somewhat disagree), but showed a lot of variance. The first statement on deception and manipulation had two modes (2 – disagree and 5 – somewhat agree), showing the divide in opinions. A comparison of the results per choice (stop/continue) can be found in Section 6.3.2.

The statements on trustworthiness and benevolence (statements 3 and 4) had medians of 5 (somewhat agree), but also a similar divide in opinions with statement 3 having two modes (3 – somewhat disagree and 5 – somewhat agree).

The last statement on user acceptance had a median of 5 (somewhat agree) and a mode of 6 (agree). This shows moderate approval for the used nudges, but a certain level of suspicion towards the system itself. A visual overview of the results per statement can be found in Figure 54, Appendix.

Statement	Median	Modes	Mean	SD
1) I think this system is deceptive or manipulative	4	2, 5	3.64	1.75
2) I am suspicious of this system's intents, actions, or outputs	3	5	3.56	1.78
3) I can trust this system	5	3, 5	4.4	1.41
4) I think this system has my best interest in mind	5	6	4.6	1.22
5)I think the applied strategies (nudges) are acceptable in this specific use case	5	6	4.72	1.56

(7-pt Likert Scale, 1=Strongly disagree, 7=Strongly agree)

TABLE 6.6: Descriptive Statistics of User Acceptance (Covert Nudges)

As for the option open-text feedback, 10 participants had additional feedback on the user acceptance. The most frequently mentioned points of concern were the default nudge being perceived as manipulative, and some of the font-choices like boldface being perceived as questionable or manipulative. One participant remarked that they felt like the provided data was insufficient to make a well-reasoned choice, which felt manipulative. Positive remarks were added about the choice of colors, the neutral wording, and the balanced presentation of the decision data. Two quotes from the open feedback are listed below, which also show the divide in user trust and acceptance between participants, especially regarding concerns about the benevolence of the system.

> P4: *"I did not feel manipulated by a pre-selected option since it is the left option, and that one makes sense to look at first. As mentioned before, the colors of the tags drew my attention, but I did not find the information very interesting for my decision, so it did not feel as a nudge. The wording feels neutral and not leaning towards a certain decision. It's good that you need to provide an explanation, this can also help in ensuring that people have read the information instead of just glancing over, especially if they come in with a biased opinion/already leaning towards a certain answer"*

Statement	Median	Modes	Mean	SD
1) I think this system is deceptive or manipulative	3	2	3.36	1.57
2) I am suspicious of this system's intents, actions, or outputs	3	3	3.2	1.58
3) I can trust this system	5	4, 6	4.88	1.16
4) I think this system has my best interest in mind	5	4, 5, 6	4.96	1.20
5)I think the applied strategies (nudges) are acceptable in this specific use case	6	6	5.2	1.25

(7-pt Likert Scale, 1=Strongly disagree, 7=Strongly agree)

TABLE 6.7: Descriptive Statistics of User Acceptance (Transparent Nudges)

P18: *"I think it would be great to have a system like this, however I would be afraid of being manipulated in a certain direction because of the nudges or that the person that sets it up could manipulate depending on what they feel is good"*

User Acceptance of Transparent Nudges

The section on the transparent nudges show slightly higher results in regard to user trust and acceptance compared to the previous covert nudges.

The first two statements on perceived manipulation and suspicion had medians of 3 (somewhat disagree) and mode values of 2 (disagree) and 3 (somewhat disagree) respectively. This showed a slight decrease in perceived manipulation, compared to the covert nudges.

The statements on trust and benevolence (statements 3 and 4) had median values of 5 (somewhat agree). Statement 3 (trust) has the modes 4 and 6 (neutral and agree), and statement 4 (benevolence) modes of 4, 5, and 6 (neutral – agree). This shows a slight increase in user trust and the perceived benevolence of the system, compared to the covert nudges.

The last statement on the user acceptance of the transparent nudges has a median and mode of 6 (agree), with a lower standard deviation compared to the design with covert nudges, showing slight improvement in terms of user acceptance. A visual overview of the results per statement can be found in Figure 55, Appendix.

As for the optional feedback, 9 participants left additional remarks. The most common feedback was that the added information cues were useful, but still lacking context and transparency (e.g. who wrote the explanations). In addition to that, the use of defaults was still remarked on as manipulative, with two participants mentioning that the added transparency would intensify the nudge and perceived manipulation (see quotes below).

> P5: *"My problem is still the pre-selection of the option. When this happens, the decision maker is given an easy path because the computer has decided. I don't think that's ideal."*

> P4: *"I feel like the explanation for the pre-selection makes it a lot more biased, i did not think that the system nudged me in that way before. The system does not feel neutral anymore"*

> P19: *"[...] Pre-selecting and even stating that it is the most rational choice is in my opinion incredibly influential. I think that offering decision support in interpreting the trade-offs between the different options is really helpful for inexperienced decision makers. But if such strong suggestions are given, they should also be defended more rigorously and transparently in my opinion."*

A direct comparison of the results per statement and design (covert/transparent) can be seen in Figure 33.

Comparison and Statistical Significance

In order to compare the results of the covert and transparent system designs and to check the statistical significance of the results, a statistical analysis was performed. First, the normality of the results was tested using the Shapiro-Wilk test. Since the research uses a within-subject design to compare two versions of the same system, the results were then checked for significance using a paired samples t-test for the normally distributed results, and the Wilcoxon Signed-Rank test for not normally distributed results.

The analysis was done using the SciPy library for Python. The results of the statistical analysis can be found in Table 6.9. The chosen significance level for the normal distribution and the statistical significance was 0.05.

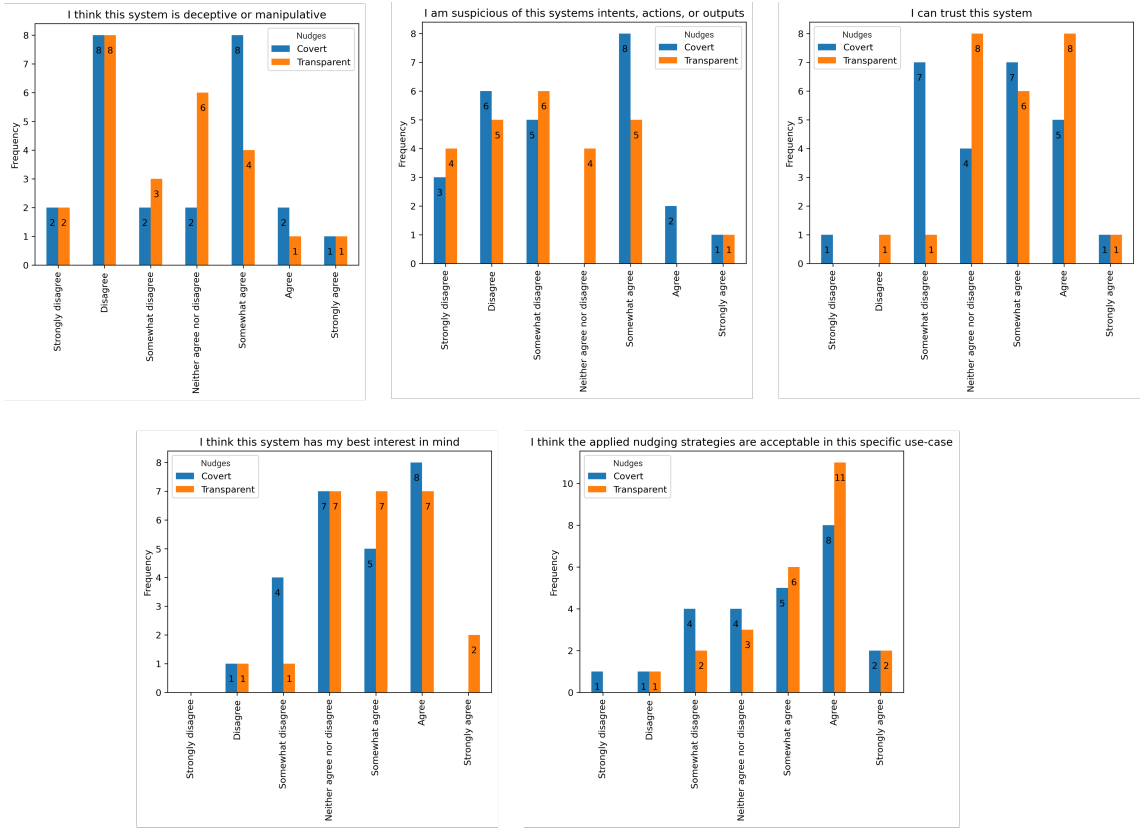

FIGURE 33: Comparison User Acceptance

Of the five compared statements, only two result sets (1 and 3) were normally distributed according to the Shapiro-Wilk test, with a p-value of <0.05. The remaining three result sets (statements 2, 4, 5) deviated from the normal distribution with a p-value above 0.05.

For the normally distributed results (1, 3), the statistical significance was then checked using a paired t-test, which resulted in a p-value > 0.05 for both statements, meaning the results are not statistically significant.

For the three statements that were not normally distributed, the significance was checked using the Wilcoxon Signed-Rank test. Of those three statements (2, 4, 5), only statement 4 (I think this system has my best interest in mind) had statistically significant results with a p-value < 0.05.

Summarizing, while the overall ratings of user acceptance for the transparent system design were slightly higher and showed lower variability than the ones for the covert design, only the statement on the system benevolence (statement 4) has statistically significant results.

In addition to that, it was analyzed whether the experience level of the participants had any influence on the results and ratings. For this, the results were divided into two groups and then compared using the Wilcoxon rank-sum test: low experience (little or no experience, and beginner), and high experience (intermediate and professional). The detailed results of this analysis can be found in Tables 1 to 3, Appendix. However, the results showed no significant differences between the two groups.

Similarly, the results were also analyzed based on the project choice made by the participants (stop or continue). Here, the results were grouped based on the choice, before being checked for significance using the Wilcoxon rank-sum test. The detailed results of this analysis can be found in Tables 4 to 6, Appendix. Only one statement in this analysis (I am suspicious of this system's intents, actions, or outputs [Transparent Design]) had statistically significant results (p-value < 0.05), which can be found in Table 6.8. The statement shows that the group that chose to stop the project rated their suspicion of the transparent system design as higher (mean 3.5) than the group that chose to continue the project (mean 2.0). This indicates that the group that chose to continue the project (biased choice) perceived the transparency cues as more trustworthy than the group that chose to stop the project.

Statement	**Decision Type**	**Mean**	**SD**	**Mann-Whitney U**	**p-value**
I am suspicious of this system's intents, actions, or outputs (Transparent Design)	Stop Project	3.50	1.61	79.00	0.0485
	Continue Project	2.00	0.71		

Table 6.8: Suspicion of System per Project Choice

Statement	**Test**	**Test Statistic**	**p-value**	**Significance**
1. I think this system is deceptive or manipulative	Paired t-test	0.825	0.417	Not Significant
2. I am suspicious of this system's intents, actions, or outputs	Wilcoxon Signed-Rank Test	48.000	0.296	Not Significant
3. I can trust this system	Paired t-test	-1.627	0.117	Not Significant
4. I think this system has my best interest in mind	Wilcoxon Signed-Rank Test	10.000	0.029	Significant
5. I think the applied strategies are acceptable in this specific use-case	Wilcoxon Signed-Rank Test	21.500	0.085	Not Significant

Table 6.9: Statistical Results per Statement

6.3.3 Threats to Validity

Possible threats to the validity of the questionnaire include the usability, target group, and sample size. Similarly to the in-person experiments, the system is tested using fictional data and evaluated using a relatively small group of participants that have limited experience with corporate project management. This likely impacted the quality and the statistical significance of the results.

Furthermore, the online questionnaire could only provide static images of the used decision support system, which led to certain functionalities being unusable (like hovering, tooltips, and other interface interactions), which might have impacted the perceived usefulness and usability. For future work, testing with working professionals and a larger sample size might produce more significant results.

Chapter 7

Discussion

This chapter will discuss the key findings and their implications, as well as the limitations of this research, and give possible directions for future work. The main two research questions, including the related assumptions and hypothesis that were explored and tested in this thesis can be found below.

RQ 1.1 *How to visualize past decisions and current project data in a project management context to help create bias awareness?*

> **Assumption** *Using an informative dashboard prior to making a decision using a decision support screen creates bias awareness and a motivation to correct biases.*
>
> **Assumption** *Transparent nudges and bias feedback integrated into a decision support screen can be used to guide users towards rational and reflected decisions while preserving free choice*

RQ 1.2 *How does the transparency of nudging in decision support tools affect the decision making process and user acceptance in project management contexts?*

> **Hypothesis a)** *Transparent nudges have a higher level of user acceptance than covert nudges in project management decision support scenarios.*

7.1 Bias Awareness

The findings from the in-person user tests indicate that while the participants exhibited some degree of bias awareness during and after using the system, it was not effective in prompting the biased participants to reconsider their choices.

All five participants of the user tests stated that they consider bias awareness to be important, and showed some degree of bias awareness while or after using the dashboard, for example through critical remarks about their own choices. However, only the three participants that chose to stop the project (rational choice) stated that the use of the dashboard had an influence on their decision making process. This could either indicate that the use of the dashboard did support bias awareness and debiasing for those participants, or that the participants had risk-averse tendencies to begin with, and were then supported in their preference through the use of the dashboard. This could be further investigated through more elaborate user tests, for example by including an assessment of personal risk-preference to the tests.

The two participants that chose the biased option also showed some level of awareness, one by stating that they selected the risky choice, and the other one stating that they know that their choice is biased. In those two cases, the debiasing failed during the feedback stage, where the participants received detailed feedback on their choice and why it might be biased. This will be discussed in more detail in the following section on covert and transparent nudging.

In one case, the extensive information on biases and their outcomes in the dashboard led to a decrease in user acceptance, as the information was perceived as confrontational and condescending. This could either be related to personal preferences, or the fact that this participant selected the biased option and received feedback from the system, which might have triggered a counter-effects, such as defiance (see next Section).

Overall, the results indicate that using the dashboard previous to making a decision might have supported bias awareness for the participants, but did not create a motivation to correct those biases in the risk-seeking participants. This shows a failure of debiasing due to lack of motivation in the risk-seeking individuals, which is the second step needed for successful debiasing [14, 68].

7.2 Transparent and Covert Nudges

As for the results of the online questionnaire, the transparent system design received a slightly higher rating in system benevolence compared to a covert design, while there were no significant differences found for the user trust or perceived manipulation.

The transparent system design did show slight improvements across all three categories of perceived manipulation and user trust, system benevolence, and nudge acceptance, compared to the covert design, as expected. However, only the results on the system benevolence were statistically significant. This only partially supports hypothesis a), showing that transparent nudges are perceived more favorably in terms of user interests. The statements on manipulation (1 and 2), user trust (3) and acceptance (5) did not show significant differences between the transparent and covert design. This could either indicate that there is no significant difference in the perception of these two system designs, or that more thorough testing with a larger sample size is needed for more significant results. Furthermore, the statistical analysis did also not show any significant differences in ratings between participants with higher levels of experience, compared to participants with lower levels of experience. This might be due to the small sample size (9 participant for low experience, 16 for high experience). More representative results about this might be gathered by testing the system on a larger scale.

For the group that chose to continue with the project (biased choice), the suspicion of the transparent system design was significantly lower than for the group that chose to stop the project, which could indicate that the transparency cues lead to increased trust in this case. This could mean that biased decision makers might especially appreciate transparency and added information related to their choices in such cases, while risk averse decision makers do not take these cues into consideration as much.

In contrast to the higher ratings of system benevolence, some participants of the online questionnaire stated that they perceived the transparency cues as more manipulative than the nudges themselves, and that this impacted the user acceptance negatively. A similar reaction was seen in the user tests, where the participants that noticed the transparency cues and nudges became increasingly suspicious of the system. Both of the biased participants already showed some scepticism towards the tool previously, but especially after noticing the transparency cues or nudges. This escalated when the system provided them with the

bias feedback, with one participant refusing to relate the feedback to their choice, and the other one having a defiant reaction to the feedback.

These differences between the questionnaire results and the in-person testing might be due to the fact that the in-person testing included the feedback-step for the biased participants, while the online questionnaire left that step out in order to compare the two system designs. While most research states that transparency does not impact the effectiveness of nudges [45, 10], some argue that strong nudging interventions might trigger counter-effects (reactance) such as defiance, especially if a nudge is perceived as intrusive, which might have happened with the feedback step during the in person tests [58, 27]. It might therefore be more promising to focus on bias awareness without the use of active interventions (like the feedback) for future work, for example by using a more thoroughly connected tool (dashboard/DSS) that openly provides decision support without the use of direct feedback when making new decisions. In addition to this, the failed debiasing might also be influenced by the level of user trust in the system. Even though the transparent design was rated slightly more positively than the covert one, the overall trust in the system was not exceptionally high to begin with. Since benevolent, transparent nudges are generally more favourably received [10], the distrust of the used transparency cues in combination with the moderate perception of the system benevolence might have negatively influenced the effectiveness of the debiasing strategies.

Regarding the transparent nudges, the transparency cues in the decision support system proved rather ineffective overall, with only two of the five participants noticing the cues at all during the in-person user tests. Therefore the nudges were covert for most of the participants during the in person user tests, which might have impacted the results. This could be due to the design of these cues, as they require user interaction such as hovering or clicking to show the information. In future research, this could be improved by offering the information openly in the interface, without the need to hover or click any components or buttons. However, since the transparent design did not show significant improvements in terms of the overall user acceptance compared to the covert design, it might also be a promising approach to either remove the cues completely, or to leave them as-is (covert for most users).

Contrary to expectations, the feedback step did also not prompt reflection for the risk-seeking participants. Instead, it seemed to reinforce their biased choice, either through misunderstandings or through defiance. While this shows a failure of the applied debiasing

strategies for some people, it also matches the findings shown in other research that people are often unwilling to accept decision aid or to reconsider their past choices, even when they are made directly aware of such biases [4, 40]. This is especially true when there is little intrinsic motivation [14, 4], which applies to this specific context, as the user test participants did not have any professional or personal stake in the decision. These results also coincide with the findings of Ramos [12], that the majority of users who receive personalized bias feedback are not willing to change their answer, or do not connect the feedback to their own behaviour.

Overall, the research results only partially support hypothesis a), with the questionnaire results showing a significant difference in the perception of the system benevolence for the transparent design, but no significant differences in the other dimensions on manipulation, trust, and acceptance. To confirm or refute the hypothesis, further testing with a larger sample size would be necessary, as the current results were inconclusive.

7.3 Data Quality and System Acceptance

As the usability, usefulness, and the quality of data are key factors when it comes to the acceptance and effectiveness of dashboards and decision support screens [30, 57], the participants of the in-person tests and the online questionnaire were asked to rate these factors. Overall, the usability, usefulness, and data relevancy of the tool were rated positively, while the data sufficiency was rated more negatively.

The usability and perceived usefulness of the system were rated positively, across both the questionnaire and the user tests, which are the main two factors related to the user acceptance of dashboards, as mentioned by Janes et al. [30]. The participants of the user test were able to use the dashboard to inform themselves about the impact and outcome of biased decisions, and to combine data from different dashboard components to gain a better overview and to find trends in the data. The main point of frustration in using the dashboard was the timeline component, which proved too complex. This could either be due to the required level of visualization literacy and the use of non-professional participants, or the heavy reliance on user interaction with this component, since it applies a pull rather than a push approach.

As for the decision support screen, the data relevancy was rated positively in the online questionnaire, while the sufficiency of the provided data was rated slightly negatively. This

matches the feedback from the user tests, that the provided data is sufficient for such simplified scenarios, but would not be enough to match more complex real life scenarios. Especially the lack of transparency regarding the calculations and estimations, as well as the human factors (e.g. affected employees) were frequently mentioned. This is one of the limitations of the research, as no actual company data or real bias estimations were used during the user tests. Besides that, the usability and ease of use of the decision support screen were rated positively by the questionnaire and user test participants, with the questionnaire results indicating that participants would find the system useful for project management scenarios.

7.4 Limitations

The main limitations of the system and the research are related to the system functionalities in itself. Due to the requirements of the system (as it is meant to be integrated with other components developed in the research group) only binary go/kill decisions can be considered, which is likely not enough to adequately represent real life scenarios. External factors that would usually be taken into account when making such project decisions, like long standing contracts (how likely is it, that a project can just be canceled without breaking any commitments or contracts?) or the affected employees, can not be considered with the current system. In realistic scenarios that would probably result in more nuanced decision options, rather than a binary decision problem with clear rational bounds.

Another limitation of this research is the used data and relatively small sample size for validation. Due to the lack of industry data, the system was tested using fictional data, making the provided scenario highly simplified. In addition to that, the system was tested using a rather small sample of participants with limited project management experience and financial literacy, which limits the validity. Testing the system with more participants with stronger professional backgrounds and more realistic company data might influence the results, especially the perception on the usefulness of such a tool. Also, doing a case study with professional project managers and real project data would likely lead to more meaningful and significant results. Using actual project data would also allow for a clearer communication with the user about the underlying data, calculations, and human factors (employees affected by choice) which was a frequently mentioned point of criticism with the current system and the fictional data.

Chapter 8

Conclusion

This thesis aimed at exploring the effectiveness of using a dashboard in combination with a decision support screen in order to support bias awareness and rational decision making in project management scenarios. For this, the system design was based on literature findings about effective dashboard design, choice architecture, and debiasing.

The main objectives of the research were evaluating whether using the system creates bias awareness in the users, and how the use of nudges impacts the user acceptance and effectiveness of the system, especially regarding transparent nudges. This was done using preliminary user tests to ensure the usability of the system, followed by a validation consisting of a questionnaire on the user acceptance of two different system designs, and in-depth qualitative user tests and interviews.

The main research question of the thesis was: *RQ: How to raise bias awareness for risk related biases in project management contexts*? The two resulting sub-questions and the related hypothesis and assumptions can be found below.

RQ 1.1 *How to visualize past decisions and current project data in a project management context to help create bias awareness?*

> **Assumption** *Using an informative dashboard prior to making a decision using a decision support screen creates bias awareness and a motivation to correct biases.*
>
> **Assumption** *Transparent nudges and bias feedback integrated into a decision support screen can be used to guide users towards rational and reflected decisions while preserving free choice*

RQ 1.2 *How does the transparency of nudging in decision support tools affect the decision making process and user acceptance in project management contexts?*

> **Hypothesis a)** *Transparent nudges have a higher level of user acceptance than covert nudges in project management decision support scenarios.*

Regarding the first research question (RQ 1.1), the results of the user tests indicate that the approach of using a dashboard in combination with a decision support system was partially successful for supporting bias awareness and providing general decision support, but had limited success in prompting reflection. Although all participants of the user tests showed some level of bias awareness, the system did not seem to motivate the participants with risk-seeking tendencies to reconsider their choices, even when faced with direct feedback. Overall, the results show that the use of dashboards can be a promising approach for creating bias awareness in project management scenarios, even though the system had limited success in debiasing risk-seeking participants.

As for the second research question (RQ 1.2), the results only partially support hypothesis a), that transparent nudges have an increased user acceptance compared to a covert system design. While the design with transparent nudges was generally rated as slightly more trustworthy and reached higher levels of user acceptance, the differences were not statistically significant except for the perception of system benevolence. Especially the participants that exhibited risk-seeking behaviour during the user tests showed an increased distrust towards the transparent nudges, and rejected them when provided with feedback. While this matches the findings from the literature that people tend to reject critical feedback related to their choices, the divide in opinions also shows the need for further testing to investigate these findings more thoroughly.

An overview of the results and the exploratory findings can be found below.

> **Hypothesis a) only partially supported** *The results of the questionnaire indicate slightly higher levels of user trust and acceptance for the transparent design, however only the statement on system benevolence showed statistically significant differences.*
>
> **Assumption on bias awareness** *The use of the dashboard previous to making a decision seemed to support bias awareness for the participants, but failed to motivate the biased decision makers to reconsider their choices.*

Assumption on transparent nudges *The transparent nudges, and the feedback step in particular, did not seem to prompt a reflection on biased decisions during the user tests.*

Overall, these results show how challenging the debiasing of decisions can be, especially in more complex scenarios like project management, and that people often tend to reject feedback related to their own choices. The results show that risk-seeking tendencies are also present among users with moderate levels of project management experience, showing the relevancy of the topic and the need for tools that can support rational choice and reflection.

The findings of this thesis also provide several directions for potential future research, such as additional validation and alternative system designs. For future work it might be interesting to see whether a more connected system design (one combined tool instead of two) would be more effective compared to the approach used in this study, as the disconnect between the two components was frequently mentioned during the user tests. For example, the system could be restructured in a way that combines the dashboard and support screen into a single tool that gives clear and transparent recommendations as to which option is rational and why, without the use of active feedback. This could also lessen the scepticism and reactance to the feedback step that was seen during the user tests. Additionally, it would be interesting to investigate whether the personal risk preferences of the participants influence the results. Understanding whether risk-averse or risk-seeking tendencies impact the effectiveness of the chosen nudges and feedback interventions would allow for more personalized decision support and debiasing. As for future validation, testing the system on a larger scale with working professionals and actual company data could provide more meaningful results and insights.

In conclusion, this thesis has explored the combination of a dashboard and a decision support screen in order to create bias awareness and to debias financial project decisions. The results show that while such a system can be a promising approach for enhancing bias awareness in project management scenarios, it had limited success for debiasing and prompting a reconsideration of choices among biased participants. These results highlight both the potential and the challenges related to debiasing and decision support in the financial domain, especially when it comes to creating bias awareness.

Bibliography

[1] Khushbu Agrawal. "A Conceptual Framework of Behavioral Biases in Finance." In: *IUP Journal of Behavioral Finance* (2012).

[2] Michael K Allio. "Strategic dashboards: designing and deploying them to improve implementation". In: *Strategy & Leadership* 40.5 (2012), pp. 24–31.

[3] Sohrab Almasi et al. "Usability evaluation of dashboards: A systematic literature review of tools". In: *BioMed Research International* 2023.1 (2023), p. 9990933.

[4] Hal R Arkes. "Costs and benefits of judgment errors: Implications for debiasing." In: *Psychological bulletin* 110.3 (1991), p. 486.

[5] Benjamin Bach et al. "Dashboard design patterns". In: *IEEE transactions on visualization and computer graphics* 29.1 (2022), pp. 342–352.

[6] G Bhandari and K Hassanein. "An agent-based debiasing framework for investment decision-support systems". English. In: *Behaviour and Information Technology* 31.5 (2012), pp. 495–507. ISSN: 0144929X. DOI: 10.1080/0144929X.2010.499477.

[7] Virginia Braun and Victoria Clarke. "Using thematic analysis in psychology". In: *Qualitative research in psychology* 3.2 (2006), pp. 77–101.

[8] John Brooke et al. "SUS-A quick and dirty usability scale". In: *Usability evaluation in industry* 189.194 (1996), pp. 4–7.

[9] Matthew Brzowski and Dan Nathan-Roberts. "Trust measurement in human–automation interaction: A systematic review". In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Vol. 63. 1. SAGE Publications Sage CA: Los Angeles, CA. 2019, pp. 1595–1599.

[10] Luca Congiu and Ivan Moscati. "A review of nudges: Definitions, justifications, effectiveness". In: *Journal of Economic Surveys* 36.1 (2022), pp. 188–213. DOI: https:

//doi.org/10.1111/joes.12453. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/joes.12453.

[11] Pierre Cossette. "Heuristics and cognitive biases in entrepreneurs: a review of the research". In: *Journal of Small Business & Entrepreneurship* 27.5 (2014), pp. 471–496.

[12] Eduardo da Costa Ramos. "ABI APPROACH: AN ONTOLOGY-BASED APPROACH FOR AUTOMATIC BIAS IDENTIFICATION IN DECISION-MAKING UNDER RISK". PhD thesis. Rio de Janeiro, BR: Universidade Federal do Rio de Janeiro, June 2024.

[13] Eduardo da Costa Ramos et al. "Extending the Core Ontology on Decision Making according to Behavioral Economics". In: *14th Ontology Research Seminar in Brazil, ONTOBRAS 2021*. CEUR. 2021.

[14] Pat Croskerry, Geeta Singhal, and Silvia Mamede. "Cognitive debiasing 1: origins of bias and theory of debiasing". In: *BMJ quality & safety* 22.Suppl 2 (2013), pp. ii58–ii64.

[15] Pat Croskerry, Geeta Singhal, and Silvia Mamede. "Cognitive debiasing 2: impediments to and strategies for change". In: *BMJ quality & safety* 22.Suppl 2 (2013), pp. ii65–ii72.

[16] J. Da. Silva Lopes, J.L. Braga, and M.A. Resende Filho. "Systems dynamics model for decision support in risk assessment in software projects". In: *Journal of Software: Evolution and Process* 27.12 (2015), pp. 976–989. DOI: 10.1002/smr.1754.

[17] Fred D Davis, RP Bagozzi, and PR Warshaw. "Technology acceptance model". In: *J Manag Sci* 35.8 (1989), pp. 982–1003.

[18] Irina Dolgopolova, Bingqing Li, and Jutta Roosen. "Do Nudges Matter? Consumer Perception and Acceptance of Recommender Systems with Different Types of Nudges". In: *3rd FAccTRec Workshop: Responsible Recommendation at The ACM Conference Series on Recommender Systems RECSYS 2020*. 2020.

[19] Doaa Farouk Badawy Eldesouky. "Visual hierarchy and mind motion in advertising design". In: *Journal of Arts and Humanities* 2.2 (2013), pp. 148–162.

[20] Geoffrey Ellis. "So, What Are Cognitive Biases?" In: *Cognitive Biases in Visualizations*. Ed. by Geoffrey Ellis. Cham: Springer International Publishing, 2018, pp. 1–10. ISBN: 978-3-319-95831-6. DOI: 10.1007/978-3-319-95831-6_1.

[21] Stephen Few. *Information dashboard design: The effective visual communication of data*. O'Reilly Media, Inc., 2006.

[22] Kraig Finstad. "The usability metric for user experience". In: *Interacting with computers* 22.5 (2010), pp. 323–327.

[23] Baruch Fischoff. "Debiasing". In: *Judgment under uncertainty: Heuristics and biases* 31 (1982).

[24] Bent Flyvbjerg. "Top ten behavioral biases in project management: An overview". In: *Project Management Journal* 52.6 (2021), pp. 531–546.

[25] Kathy Haan. *Best Free Project Management Software (2023) — forbes.com*. `https://www.forbes.com/advisor/business/software/free-project-management-software/`. [Accessed 12-May-2023].

[26] F. Haghighi Rad and S.M. Rowzan. "Designing a hybrid system dynamic model for analyzing the impact of strategic alignment on project portfolio selection". In: *Simulation Modelling Practice and Theory* 89 (2018), pp. 175–194. DOI: `10.1016/j.simpat.2018.10.001`.

[27] William Hagman. "When are nudges acceptable?: Influences of beneficiaries, techniques, alternatives and choice architects". PhD thesis. Linköping University Electronic Press, 2018.

[28] John S Hammond, Ralph L Keeney, and Howard Raiffa. "The hidden traps in decision making". In: *Harvard business review* 76.5 (1998), pp. 47–58.

[29] Ralph Hertwig and Till Grüne-Yanoff. "Nudging and boosting: Steering or empowering good decisions". In: *Perspectives on Psychological Science* 12.6 (2017), pp. 973–986.

[30] Andrea Janes, Alberto Sillitti, Giancarlo Succi, et al. "Effective dashboard design". In: *Cutter IT journal* 26.1 (2013), pp. 17–24.

[31] Jiun-Yin Jian, Ann M Bisantz, and Colin G Drury. "Foundations for an empirically determined scale of trust in automated systems". In: *International journal of cognitive ergonomics* 4.1 (2000), pp. 53–71.

[32] Eric J Johnson et al. "Beyond nudges: Tools of a choice architecture". In: *Marketing letters* 23 (2012), pp. 487–504.

[33] Daniel Kahneman. "Maps of bounded rationality: Psychology for behavioral economics". In: *American economic review* 93.5 (2003), pp. 1449–1475.

[34] Daniel Kahneman and Shane Frederick. "A Model of Heuristic Judgment". In: *The Cambridge Handbook of Thinking and Reasoning*. Ed. by K. Holyoak and B. Morrison. Cambridge University Press, 2005, pp. 267–293.

[35] Daniel Kahneman and Amos Tversky. "Prospect theory: An analysis of decision under risk". In: *Handbook of the fundamentals of financial decision making: Part I*. World Scientific, 2013, pp. 99–127.

[36] Daniel Kahneman and Amos Tversky. "The psychology of preferences". In: *Scientific American* 246.1 (1982), pp. 160–173.

[37] Katarzyna Katner and Radu Jianu. "The effectiveness of nudging in commercial settings and impact on user trust". In: *Extended abstracts of the 2019 chi conference on human factors in computing systems*. 2019, pp. 1–6.

[38] Mark Keil, Gordon Depledge, and Arun Rai. "Escalation: The role of problem recognition and cognitive bias". In: *Decision Sciences* 38.3 (2007), pp. 391–421.

[39] Johan E Korteling and Alexander Toet. "Cognitive biases". In: *Encyclopedia of Behavioral Neuroscience,* (2022), pp. 610–619.

[40] Richard P Larrick. "Debiasing". In: *Blackwell handbook of judgment and decision making* (2004), pp. 316–338.

[41] Jessica Nina Lester, Yonjoo Cho, and Chad R Lochmiller. "Learning to do qualitative data analysis: A starting point". In: *Human resource development review* 19.1 (2020), pp. 94–106.

[42] James R Lewis. "IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use". In: *International Journal of Human-Computer Interaction* 7.1 (1995), pp. 57–78.

[43] Myles E Mangram. "A simplified perspective of the Markowitz portfolio theory". In: *Global journal of business research* 7.1 (2013), pp. 59–70.

[44] Nuno Martins, Susana Martins, and Daniel Brandão. "Design principles in the development of dashboards for business management". In: *Perspectives on Design II: Research, Education and Practice* (2022), pp. 353–365.

[45] Craig RM McKenzie et al. "Constructed preferences, rationality, and choice architecture". In: *Review of Behavioral Economics* 5.3-4 (2018), pp. 337–360.

[46] Christian Meske and Tobias Potthoff. “The DINU-model–a process model for the design of nudges”. In: (2017).

[47] monday.com. *Monday Apps Framework — monday.com*. https://developer.monday.com/apps/docs/intro. [Accessed 21-June-2024].

[48] monday.com. *Vibe Design system — monday.com*. https://developer.monday.com/apps/docs/vibe-design-system. [Accessed 21-June-2024].

[49] A. Poligadu and R.K. Moloo. “An innovative measurement programme for agile governance”. In: *International Journal of Agile Systems and Management* 7.1 (2014), pp. 26–60. DOI: 10.1504/IJASM.2014.059153.

[50] Ariane MB Rodrigues et al. “Comparing the effectiveness of visualizations of different data distributions”. In: *2019 32nd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*. IEEE. 2019, pp. 84–91.

[51] A.T. Sadabadi. *A semi-supervised learning framework for decision modeling of software project management*. Vol. 465. 2014, pp. 134–149. DOI: 10.1007/978-3-319-11958-8{_}11.

[52] Bahador Saket, Alex Endert, and Çağatay Demiralp. “Task-based effectiveness of basic visualizations”. In: *IEEE transactions on visualization and computer graphics* 25.7 (2018), pp. 2505–2512.

[53] Alper Sarikaya et al. “What do we talk about when we talk about dashboards?” In: *IEEE transactions on visualization and computer graphics* 25.1 (2018), pp. 682–692.

[54] Leonardo F Sayão and Fernanda A Baião. “An Ontology-based Data-driven Architecture for Analyzing Cognitive Biases in Decision-making”. In: (2023).

[55] R. M. Skitmore, S. G. Stradling, and A. P. Tuohy. “Project management under uncertainty”. In: *Construction Management and Economics* 7.2 (1989), pp. 103–113. DOI: 10.1080/01446198900000015.

[56] Paul Slovic, Baruch Fischhoff, and Sarah Lichtenstein. “Behavioral decision theory”. In: *Annual review of psychology* 28.1 (1977), pp. 1–39.

[57] Jack B Soll, Katherine L Milkman, and John W Payne. “A user’s guide to debiasing”. In: *The Wiley Blackwell handbook of judgment and decision making* 2 (2015), pp. 924–951.

[58] Cass R Sunstein. “Nudges that fail”. In: *Behavioural public policy* 1.1 (2017), pp. 4–25.

[59] Richard H Thaler and Cass R Sunstein. “Libertarian paternalism”. In: *American economic review* 93.2 (2003), pp. 175–179.

[60] Richard H Thaler and Cass R Sunstein. *Nudge: Improving decisions about health, wealth, and happiness*. Penguin, 2009.

[61] Richard H Thaler, Cass R Sunstein, and John P Balz. “Choice architecture”. In: *The behavioral foundations of public policy* (2014).

[62] de Haan Thomas and Linde Jona. “’Good Nudge Lullaby’: Choice Architecture and Default Bias Reinforcement”. In: *The Economic Journal* 128.610 (2018), pp. 1180–1206.

[63] Amos Tversky. “A critique of expected utility theory: Descriptive and normative considerations”. In: *Erkenntnis* (1975), pp. 163–173.

[64] Amos Tversky and Daniel Kahneman. “Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty.” In: *science* 185.4157 (1974), pp. 1124–1131.

[65] Amos Tversky and Daniel Kahneman. “The framing of decisions and the psychology of choice”. In: *science* 211.4481 (1981), pp. 453–458.

[66] Dante A Urbina and Alberto Ruiz-Villaverde. “A critical review of homo economicus from five approaches”. In: *American Journal of Economics and Sociology* 78.1 (2019), pp. 63–93.

[67] Elke U Weber. “Risk attitude and preference”. In: *Wiley Interdisciplinary Reviews: Cognitive Science* 1.1 (2010), pp. 79–88.

[68] Timothy D Wilson and Nancy Brekke. “Mental contamination and mental correction: unwanted influences on judgments and evaluations.” In: *Psychological bulletin* 116.1 (1994), p. 117.

[69] Ogan M Yigitbasioglu and Oana Velcu. “A review of dashboards in performance management: Implications for design and research”. In: *International Journal of Accounting Information Systems* 13.1 (2012), pp. 41–59.

Appendix

FIGURE 34: Dashboard Overview Figma

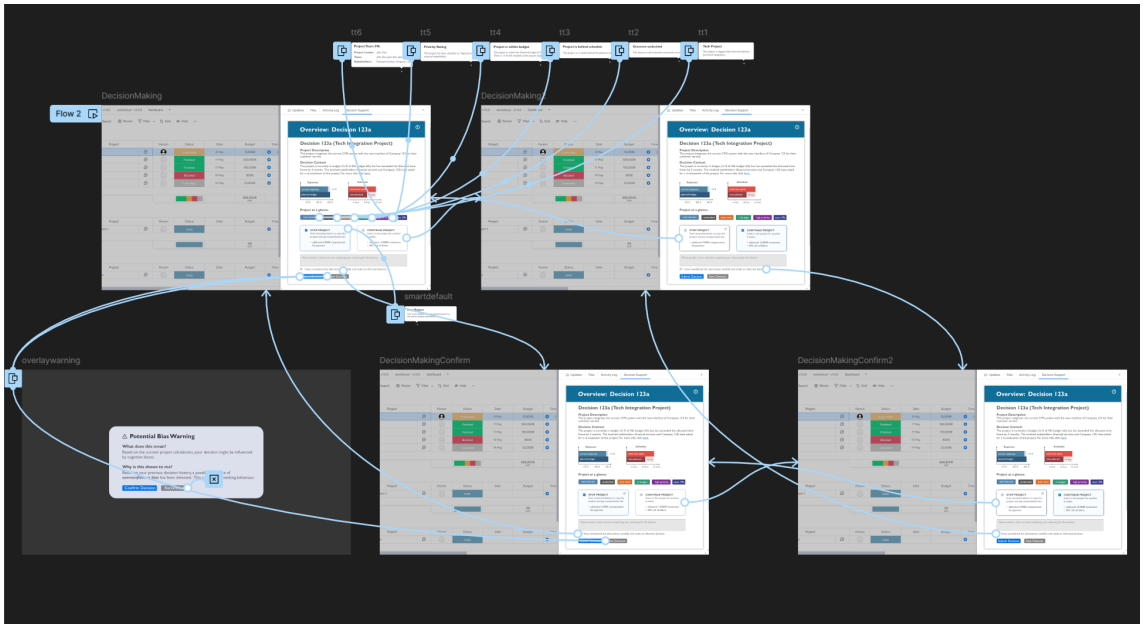

FIGURE 35: DSS Overview Figma

Consent Form for the Usability Study

(YOU WILL BE GIVEN A COPY OF THIS INFORMED CONSENT FORM)

Please tick the appropriate boxes	**Yes**	**No**
Taking part in the study		
I have read and understood the information brochure or it has been read to me.	○	○
I have been able to ask questions about the study and my questions have been answered to my satisfaction.	○	○
I consent voluntarily to be a participant in this study and understand that I can refuse to answer questions and I can withdraw from the study at any time, without having to give a reason.	○	○
I understand that taking part in the study involves a video- and audio recorded usability test of a web-based tool, which will be transcribed and used (for example screenshots of the performed actions) in a written report.	○	○
I understand that taking part in the study involves a video- and audio-recorded short interview on the usability of the presented web-based prototypes, which will be transcribed and anonymized for a written report.	○	○
Use of the information in the study		
I understand that information I provide will be used for the written report of a master thesis	○	○
I understand that personal information collected about me that can identify me, such as my name or other personal data will not be shared beyond the researchers and employees involved with this specific master thesis.	○	○
I agree that my information can be quoted in research outputs	○	○
I agree to be audio and video recorded.	○	○

Signatures

_____	_____	_____
Name of participant	Signature	Date

I have accurately read out the information sheet to the potential participant and, to the best of my ability, ensured that the participant understands to what they are freely consenting.

_____	_____	_____
Researcher name	Signature	Date

Contact details for further information: Saskia Hustinx s.c.hustinx@student.utwente.nl

UNIVERSITY OF TWENTE.

FIGURE 36: Usability Test Consent Form

Information Sheet – Usability Study Web-based Debiasing Tools:

INTRODUCTION:

I would like to invite you to take part in a user study for my master thesis. Before you decide, you need to understand why the research is being done and what it would involve for you. Please take time to read the following information carefully. Ask questions if anything you read is not clear or if you would like more information. Take time to decide whether or not to take part.

WHO I AM AND WHAT THIS STUDY IS ABOUT?

This is a study for a master thesis in Interaction-Technology (I-Tech), researching the usability and reception of different prototypes for a web-based tool. The results of the study will only be used within the context of the aforementioned thesis and related outputs, such as the thesis itself and any papers that might be published about it.

WHAT WILL TAKING PART INVOLVE?

The study will involve a usability test of different prototypes with a short interview on your opinion and experience with the prototypes. The study will be short, about 30-minutes, and take place at the university, room (TODO). The usability tests and interviews will be audio and video-recorded for further analysis and transcription.

WHY HAVE YOU BEEN INVITED TO TAKE PART?

You have been selected for this part of the study, due to your interest in the topic when you were asked to participate.

DO YOU HAVE TO TAKE PART?

The participation in this research is completely voluntary and you, as the participant, have the right to refuse participation, refuse any question and withdraw at any time without any consequence whatsoever.

WHAT ARE THE POSSIBLE RISKS AND BENEFITS OF TAKING PART?

There are no risks involved with participating in this study. The incentive for participation will be a pastry and a coffee, as mentioned when you agreed to take part in the research.

WILL TAKING PART BE CONFIDENTIAL?

The gathered data of this research will be handled confidentially and stored in the University-related storage (Onedrive). The gathered data and results will only be used within the scope of the thesis and will only be used in an anonymized way in the resulting report.

FIGURE 37: Usability Test Information Letter

PSSUQ (Post-Study System Usability Questionnaire)

7-point Likert Scale (Strongly agree to strongly disagree + NA option) for the following 16 items: (https://uiuxtrend.com/pssuq-post-study-system-usability-questionnaire/)

1. Overall, I am satisfied with how easy it is to use this system.
2. It was simple to use this system.
3. I was able to complete the tasks and scenarios quickly using this system.
4. I felt comfortable using this system.
5. It was easy to learn to use this system.
6. I believe I could become productive quickly using this system.
7. The system gave error messages that clearly told me how to fix problems.
8. Whenever I made a mistake using the system, I could recover easily and quickly.
9. The information (such as online help, on-screen messages, and other documentation) provided with this system was clear.
10. It was easy to find the information I needed.
11. The information was effective in helping me complete the tasks and scenarios.
12. The organization of information on the system screens was clear.
13. The interface of this system was pleasant.
14. I liked using the interface of this system.
15. This system has all the functions and capabilities I expect it to have.
16. Overall, I am satisfied with this system.

System usability scale (SUS)

5-point Likert Scale (Strongly agree to strongly disagree) for the following 10 items: (https://uiuxtrend.com/measuring-system-usability-scale-sus/ ; Brooke, John. "Sus: a “quick and dirty’ usability." Usability evaluation in industry 189.3 (1996): 189-194.)

1. I think that I would like to use this website frequently.
2. I found the website unnecessarily complex.
3. I thought the website was easy to use.
4. I think that I would need the support of a technical person to be able to use this website.
5. I found the various functions in this website were well integrated.
6. I thought there was too much inconsistency in this website.
7. I would imagine that most people would learn to use this website very quickly.
8. I found the website very cumbersome to use.
9. I felt very confident using the website.
10. I needed to learn a lot of things before I could get going with this system.

FIGURE 38: SUS and PSSUQ Questionnaires

Task	Participant 1 (sec)	Participant 1 Remarks	Participant 2 (sec)	Participant 2 Remarks	Participant 3 (sec)	Participant 3 Remarks	Average Time (sec)	Overall User Effectiveness
Can you tell me how many biased decisions were taken in April of 2023?	**6,40**		**11,89**	tried to read it off the graph instead of using the overlay	**14,12**	first mis-reads the tooltip info; then corrects	**10,80**	100%
Can you tell me how many of those had a negative outcome?	**0,97**		**21,49**	confusion about where to see this info; has overlay open but does not use it; asks to repeat the question	**0,87**	very quick after first mis-read in prev. task	**7,78**	100%
Can you tell me a reason, why the last decision was flagged as biased?	**7,95**	assuming the last decision is the top one in the detail view; names the biases itself, not the reasoning	**22,55**	long duration due to reading biases and full reasoning out loud;	**22,44**	slightly confused about order of items; then found quickly; reads out reasoning first	**17,65**	100%
Can you tell me a way to prevent risk seeking behavior?	**10,95**	long duration due to reading out the whole text	**25,76**	long duration due to reading text out loud; some confusion about wording/missing context	**9,76**		**15,49**	100%
Can you tell me what percentage of biased decisions have been reconsidered?	**5,76**	confusion with reconsidered vs. changed	**4,34**		**20,53**	confusion about reconsidering and changing, if changed decisions should be included in count	**10,21**	100%
Can you tell me the average financial impact of a biased decision?	**4,11**		**dnf**	could not find info; though previously read through it during exploration; tries to find it in bottom row diagrams	**6,51**		**5,31**	67%
Can you tell me the domain, in which most biased decisions happen?	**1,95**		**3,90**		**2,73**		**2,86**	100%
Can you tell me the project stage, in which most biased decisions happen?	**1,87**		**dnf**	did not find the interaction to switch between diagrams	**10,74**	some minor issues with switching the view	**6,31**	67%
Can you tell me what percent of biased decisions have a negative outcome?	**3,80**		**11,94**		**3,82**		**6,52**	100%
Effectiveness/Completion Rate per User	9/9 = **100%**		7/9 = **77%**		9/9 = **100%**		**Avg. completion rate:**	**92%**
(successfull tasks/all tasks undertaken)								

FIGURE 39: Results per Task

Notes Usability Testing: DSS Screen

- Participant 1
  - Reads Information Carefully, Top Down; uses hover options
  - Asks for some clarification on why decision must be taken. (Clarification: Stakeholders asked for re-evaluation)
  - Apparently the 'planning part' is not as obvious, some confusion on how long investment etc; then cleared up, once the option part is read through;
  - Says there must be a way to calculate best outcome with risk
  - Missing some kind of progress indicator, on what has already been completed (!) as it would influence decision
  - Hard to generalize such complex decisions: too much factors to take into account
  - Gives explanation/reasoning as prompted by interface
  - Clarify what happens in case of failure more clearly: are there more payments, risks involved etc.
  - Chooses to continue
  - When bias warning popped up: comment: 'oh, I was well aware'; already suspected bias
  - Gets warning, then confirms decision
  - Note: did not notice 'nudge info' or pre-selected default
- Participant 2
  - Reads info top down, carefully reading all parts/written info specifically
  - Does not use much of the hovering functions
  - Tries to use the links for more info
  - Uses graphs -> misreads captions at first
  - Likes 'at-a-glance' functionality; does not use hover; tries to piece together more info
  - Reads through options carefully
  - Would like more info about how much you're supposed to type into the reasoning field
  - Reasons that investment would probably not be enough; chooses 'stop project' and saves decision
  - When bias warning pops up:
    - Tries to reconsider decision
  - Did not notice default or 'nudge' info
- Participant 3
  - Briefly skips over written text
  - Bit of confusion about graphs
  - Uses hover functions
  - Chooses default
  - Submits decision
  - Bias warning:
    - Reads bias warning more carefully
    - Tries to change answer
  - Did not notice default or 'nudge' info

FIGURE 40: Notes about Decision Screen Usability Test

Usability Interview Results

Usability

What was your overall impression of the user interface?

- P1:
  - Very understandable, probably even more clear for professionals of that domain
  - Hard to generalize/condense such a decision
- P2
  - Is well organized in different blocks per topic/theme
  - Likes the colors; makes it easier to differentiate
  - Very informative, even for non professionals
- P3:
  - Dashboard very nice, gives quick overview of situation
  - Detail view had a bit too much information to quickly skim over (not initially perceived as normal dashboard item like the diagrams)
  - Once the tasks asked it, the information was easy to find in detail view as well
  - The graphs in decision making interface are hard to understand, also looks too similar to each other
  - Instinct when given the warning screen: it is a prompt to change decision

Was did you think of the organization of information on the screen?

- P1: Very clear; simplicity is nice;
- P2:
  - enough information without being overloaded;
  - sometimes missing some context on what the data is/where from?
  - Nice that it is sorted by priority/importance for attention
  - Nice that you can get more detailed info if wanted
  - Bias info block was very distinct from rest; good to differentiate
- P3:
  - Dashboard nice, titles of elements really show at a glance what the content is; logical layout; as expected

Was there anything you particularly liked/disliked?

- P1:
  - info on demand is nice if you want more info
  - Some elements pull more attention than others (especially big numbers)
- P2: see previous question
- P3: likes dashboard, dislikes graphs in decision making interface

How would you rate the complexity of the interface?

- P1: very low/easy
- P2: adequate, no information overload
- P3: dashboard very clear, decision screen a bit confusing with the graphs

Information

Did have any difficulties to find the information needed for the tasks? If so, elaborate.

- P1: no, very easy to find
- P2:
    - Difficulties with one/two tasks
    - For the rest it was easy to find after going over dashboard
    - Pretty intuitive
- P3: no real difficulties, took some time to register detail view; took some time to figure out dropdown item

Was the content of the dashboard clear and understandable?

- P1: yes, content was understandable
- P2:
    - some words were unfamiliar; maybe add some hyperlink with examples/definitions
    - nice that there were the question marks, assuming that they would give more context or info on who to contact for more info
- P3: yes, very understandable

Were there any sections of the system that you found confusing?

- P1: graphs in decision making interface were a bit confusing at first glance
- P1: graph either reconsidered/changed is confusing, especially the labels (are changed decisions not reconsidered?)
- P2: not really; some words/numbers were unfamiliar, but probably due to background as amateur; misread some things in decision making screens
- P3: not really, mostly clear. The big number items are most clear and accessible/at a glance
- P3: reconsidered/changed is confusing

Scenarios

Which tasks were easy to complete, and why?

- P1: all of them, except decision
- P2: reading diagrams/graphs

- P3: mostly the diagram ones, but also the one regarding the info box (how can I avoid risk seeking behaviour?)

Which tasks were difficult to complete, and why?

- P1: Decision scenario was harder to complete due to unfamiliarity with topic/project
- P2: more detailed info took some more time
- P3: detail view, tasks with more reading; the reconsidered/changed diagram

Did anything about the system frustrate you when you were using it?

- P1: unclarity about which parts are clickable
- P2: confusion about which functions works in prototype
- P2: maybe add some functionality to contact someone to discuss certain decisions

Overall

Which improvements would you suggest to make the dashboard more user-friendly?

- P1: not sure if the bottom row of diagrams delivers an appropriate amount of value for the space they take up
- P1: detail view and timeline could stand out a bit more; colored items take up more attention
- P2: make everything functional; every graph has different color scheme; makes it easier to differentiate; stay away from too bright colors; just missing some more context, otherwise very clean/simple
- P3: Maybe clarify detail view somehow?

Does the system have all the functions and capabilities you would expect it to have?

- P1: more than expected initially
- P2: no expertise in area
- P3: it matches the expectations; just a bit unsure on what prompts the decision making screen. Is it a timed event? Reconsider your decisions every x days?

Do you think that such a system would be useful in project management scenarios?

- P1: hard to relate to topic; questionable if people would actually use it, or if it would be like a chore
- P1: hard to generalize; don't know if managers would take well to being told what to do
- P1: more biases might increase value
- P2: can be helpful to have overview of decision history and related biases; also for current decisions it is nice to have a somewhat simplified overview
- P2: tool would probably bias decision itself, since it could be understood as prompt to change decision; Not sure if managers would take feedback well (could also push users to defy suggestions out of spite?)

- P3: The system now only takes into account money/time; isn't a decision way more complex; also social factors etc.?

Are there any additional remarks, you would like to make?

- P1: Looks clean, content is understandable
- P2: -
- P3: -

Additional Remarks/Notes taken during research

- Graphs in second interface confusing
- Add some kind of milestone/progress bar in decision screen
- Consider value of bottom row in dashboard (?)
- Tool would be more valuable if it included more biases
- Some graphs are missing the captions/handlebars
- Add prompt on how much info is needed in info box
- None of the participants noticed pre-selected default or info item for that
- Participants swerve from expected attention pattern
- Preprocessing through colors nice/noted
- Some task based issues with second participant
- Changed/reconsidered diagram creates most confusion
- Depending on preference: user either really appreciate or dislike detail view

FIGURE 41: Summary and Notes Usability Interview

Information Sheet – In-person Testing
User Testing of a Decision Support Tool:

INTRODUCTION:

I would like to invite you to take part in a user study for my master thesis. Before you decide, you need to understand why the research is being done and what it would involve for you. Please take time to read the following information carefully. Ask questions if anything you read is not clear or if you would like more information. Take time to decide whether or not to take part.

WHO I AM AND WHAT THIS STUDY IS ABOUT?

This is a study for a master thesis in Interaction-Technology (I-Tech), researching the usability, usefulness, and effectiveness of a web-based decision support tool. The results of the study will only be used within the context of the aforementioned thesis and related outputs, such as the thesis itself and any papers that might be published about it.

WHAT WILL TAKING PART INVOLVE?

The study will involve testing a dashboard and a decision support tool, as well as taking part in a short interview on your opinion and experience with the tool. The study will take about 1 hour and take place at the university. The user tests and interviews will be screen and audio-recorded for further analysis and transcription.

WHY HAVE YOU BEEN INVITED TO TAKE PART?

You have been selected for this part of the study, due to your interest in the topic when you were asked to participate.

DO YOU HAVE TO TAKE PART?

The participation in this research is completely voluntary and you, as the participant, have the right to refuse participation, refuse any question and withdraw at any time without any consequence whatsoever.

WHAT ARE THE POSSIBLE RISKS AND BENEFITS OF TAKING PART?

There are no risks involved with participating in this study. The incentive for participation will be a pastry and a coffee, as mentioned when you agreed to take part in the research.

WILL TAKING PART BE CONFIDENTIAL?

The gathered data of this research will be handled confidentially and stored in the University-related storage (Onedrive). The gathered data and results will only be used within the scope of the thesis and will only be used in an anonymized way in the resulting report.

HOW WILL INFORMATION YOU PROVIDE BE RECORDED, STORED AND PROTECTED?

The data will be stored in a university-related Onedrive-account for the duration of the thesis research. After completing the thesis research, any personally identifiable information (PII) will be deleted. An anonymized version of the data will be used in the project report, and can therefore be accessed and retained by UT staff involved with the thesis. You, as a participant, have the right to request access, rectification or erasure of personal data at any time.

WHAT WILL HAPPEN TO THE RESULTS OF THE STUDY?

The results of the study (gathered data/answers, screenshots, transcripts etc.) will be used in the master thesis report, and published on the related university website (https://essay.utwente.nl).

WHO SHOULD YOU CONTACT FOR FURTHER INFORMATION?

For any further information you can contact the researcher: Saskia Hustinx, (s.c.hustinx@student.utwente.nl)

If you have questions about your rights as a research participant, or wish to obtain information, ask questions, or discuss any concerns about this study with someone other than the researcher(s), please contact the supervisors of the thesis, r.guizzardi@utwente.nl and m.theune@utwente.nl.

Additionally, you can contact the Secretary of the Ethics Committee of the Faculty of Electrical Engineering, Mathematics and Computer Science at the University of Twente through ethicscommittee-cis@utwente.nl

Thank you for participating!

FIGURE 42: Information Letter User Testing

Consent Form for the User Study

(YOU WILL BE GIVEN A COPY OF THIS INFORMED CONSENT FORM)

Please tick the appropriate boxes	**Yes**	**No**
Taking part in the study		
I have read and understood the information brochure or it has been read to me.	○	○
I have been able to ask questions about the study and my questions have been answered to my satisfaction.	○	○
I consent voluntarily to be a participant in this study and understand that I can refuse to answer questions and I can withdraw from the study at any time, without having to give a reason.	○	○
I understand that taking part in the study involves a screen- and audio recorded user test of a web-based tool, which will be transcribed and used (for example screenshots of the performed actions) in a written report.	○	○
I understand that taking part in the study involves an audio-recorded interview on the usability, usefulness, and effectiveness of the presented tool, which will be transcribed and anonymized for a written report.	○	○
Use of the information in the study		
I understand that information I provide will be used for the written report of a master thesis	○	○
I understand that personal information collected about me that can identify me, such as my name or other personal data will not be shared beyond the researchers and employees involved with this specific master thesis.	○	○
I agree that my responses can be quoted in research outputs	○	○
I agree to be audio and screen-recorded.	○	○

Signatures

_____	_____	_____
Name of participant	Signature	Date

I have accurately read out the information sheet to the potential participant and, to the best of my ability, ensured that the participant understands to what they are freely consenting.

_____	_____	_____
Researcher name	Signature	Date

Contact details for further information: Saskia Hustinx s.c.hustinx@student.utwente.nl

UNIVERSITY OF TWENTE.

FIGURE 43: Informed Consent User Testing

Decision Scenario

Imagine you are a project manager at a mid-sized company, currently responsible for overseeing and managing several different projects. One of these projects is now up for re-evaluation due to some delays and budget concerns. You need to make a critical go/kill decision—whether to continue investing in this project or to stop it altogether.

At your disposal, you have (1) a dashboard with your past decision data and (2) a decision-support screen.

Task Instructions: Begin by examining the dashboard, which contains data on your past project decisions. Then, use the decision-support screen to review the specific details and options available before making your decision.

Note*: Please think aloud as you go through each step, verbalizing your thoughts and considerations.*

FIGURE 44: Decision Scenario User Testing

Script In-person User Testing

Category	Step	Description	Data	Duration
Introduction (~15mins)	Formalities	Go over information letter, sign consent form		<5mins
	Give Introduction	Describe the project context and explain the structure of the session		5mins
	START RECORDINGS!			
	Demographics	Collect demographic data	What is your age? How would you rate your level of experience with project management? - Little to no experience - Beginner (less than 1 year experience) - Intermediate (1-3 years experience) - Professional (3+ years experience) In what contexts do you use project management? (school, work, etc?)	<5mins
	Introduce System and Context	Introduce the participant to the system (dashboard and dss screen) and provide them with a decision scenario	Give scenario	5mins
Testing of the dashboard (~15mins)	Dashboard Exploration	Let the user go over the dashboard using think aloud	Think aloud transcripts	<5mins
	Perform Tasks	Let the user perform some simple tasks with the dashboard	Tasks -- How would you use the dashboard to: - Describe development of decision history over time - Describe the average outcome and impact of decisions - Look up last made decision and describe it in detail	5mins
	Dashboard Usability Interview	Let user very briefly reflect on dashboard usability (questions partially adapted from UMUX questionnaire)	- What is your general impression of the dashboard? - Did you find the dashboard data adequate for completing your tasks, and why/or why not?	5mins

			- Were there any points of frustration or difficulties you encountered? - Did you think the dashboard was easy to use or not? - Any other remarks?	
Testing of the Decision Support Screen (DSS)(10mins) *[using the version with transparent nudges, since they are more likely to preserve user trust]*	Exploration	Keeping the provided decision scenario in mind, ask the participants to go over the DSS using think aloud and to make their decision using the provided data (mentioning what they think, see, notice)	Think aloud transcripts	<5mins
	Decision making	Ask the participants to make the decision, using think aloud If participants choose the 'biased' option (the one that is not nudged), they will receive feedback and have the option to modify their decision if they want to. If they choose the nudges 'rational' option, they will get a message confirming their choice (that it was saved)	The selected option & think aloud transcripts	<5mins
	Follow up interview on reasoning	Ask participants what influenced their choice	- why did you choose this option, and which elements influenced you? If they got feedback: - did they reconsider? - If so, why? For both: - did you find the decision support screen helpful for making your decision, or not? - Are there any specific elements of the tool that stood out positively or negatively?	5mins
Follow up interview (~10mins)	Interview	ask the participants about usefulness and	- How would you describe your experience with the system as a whole?	10mins

(qualitative data on system as a whole)		usability of the entire system ask participant whether using the dashboard and learning about decision-history/biases influenced their decision	- How would you rate the usefulness of the system for project managers? - Are there any challenges or drawbacks that you would expect with using this tool in project management scenarios? - Were there any features or sections of the system that you found confusing or unnecessary? - Are there any additional features/functionalities that would improve the system in your opinion? - Did using the dashboard influence your decision-making process? If so, how? - Do you think it is important for project managers to be aware of their own biases when making decisions? And why/or why not? - Was there any information you learned during the use of the system, that was surprising or new to you?	
	Open remarks	Ask for any other remarks/questions that the participants might have	- Do you have any additional remarks or comments?	
	STOP RECORDINGS!			

FIGURE 45: User Test Script

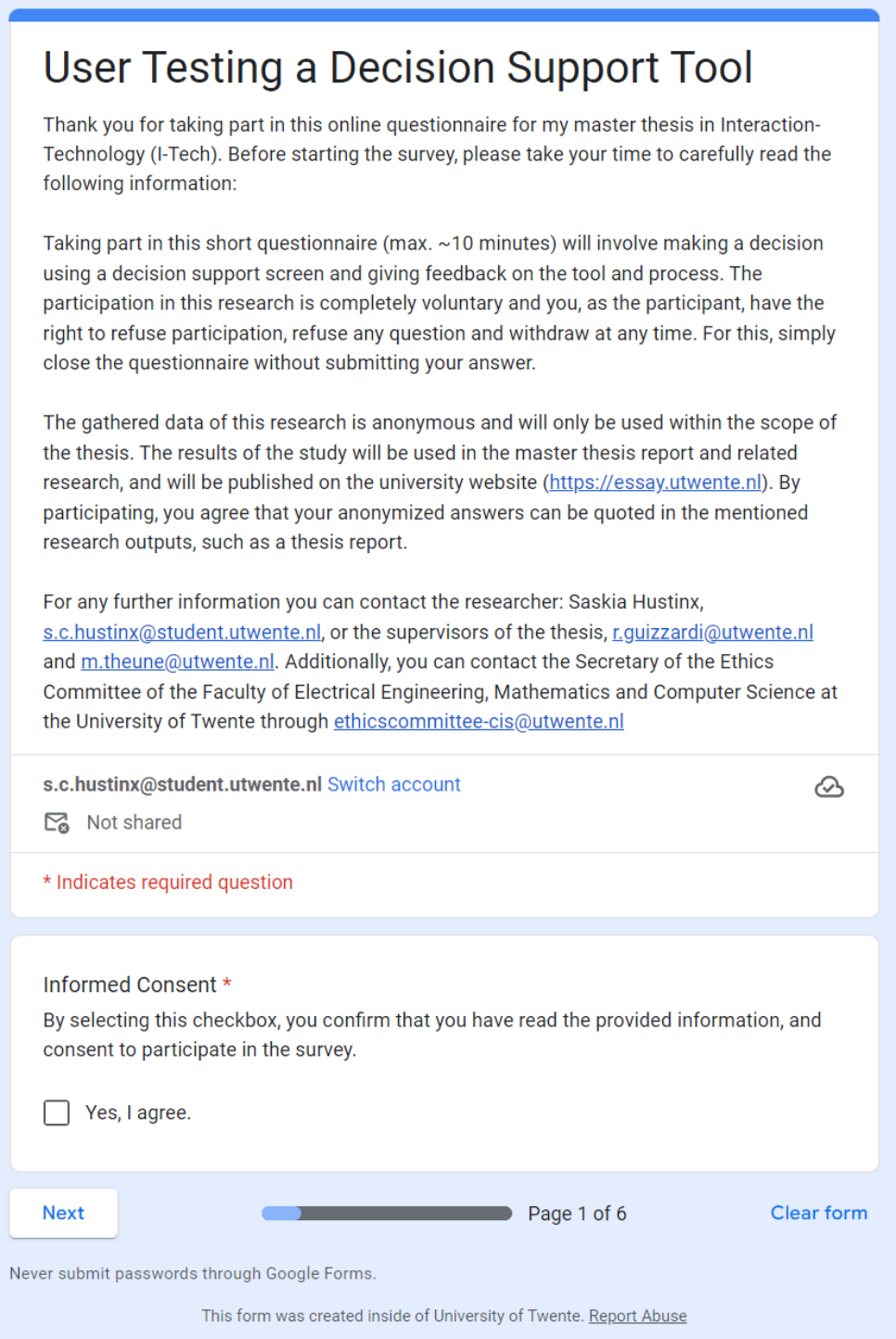

FIGURE 46: Informed Consent Online Questionnaire

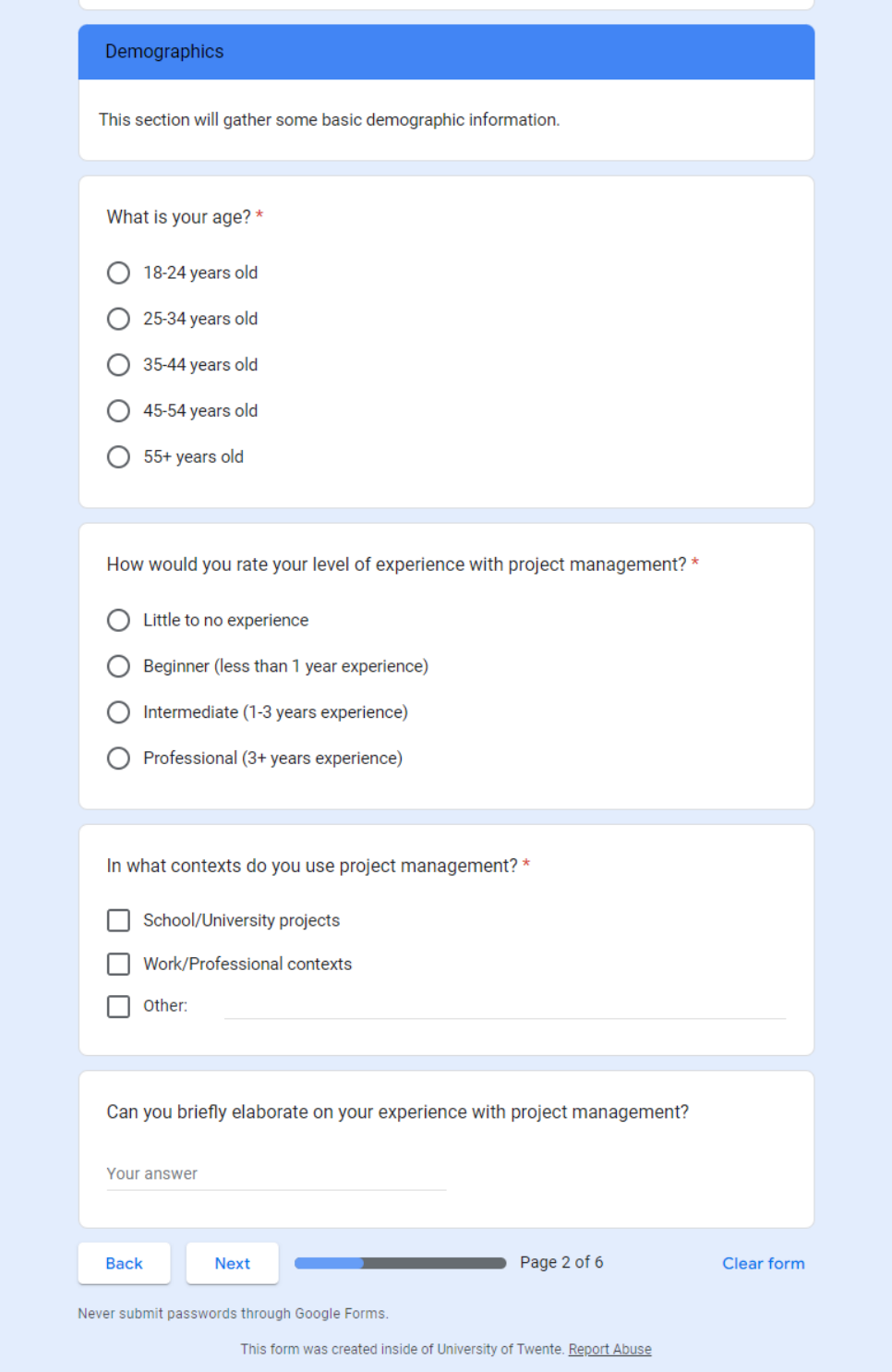

FIGURE 47: Demographic Section Questionnaire

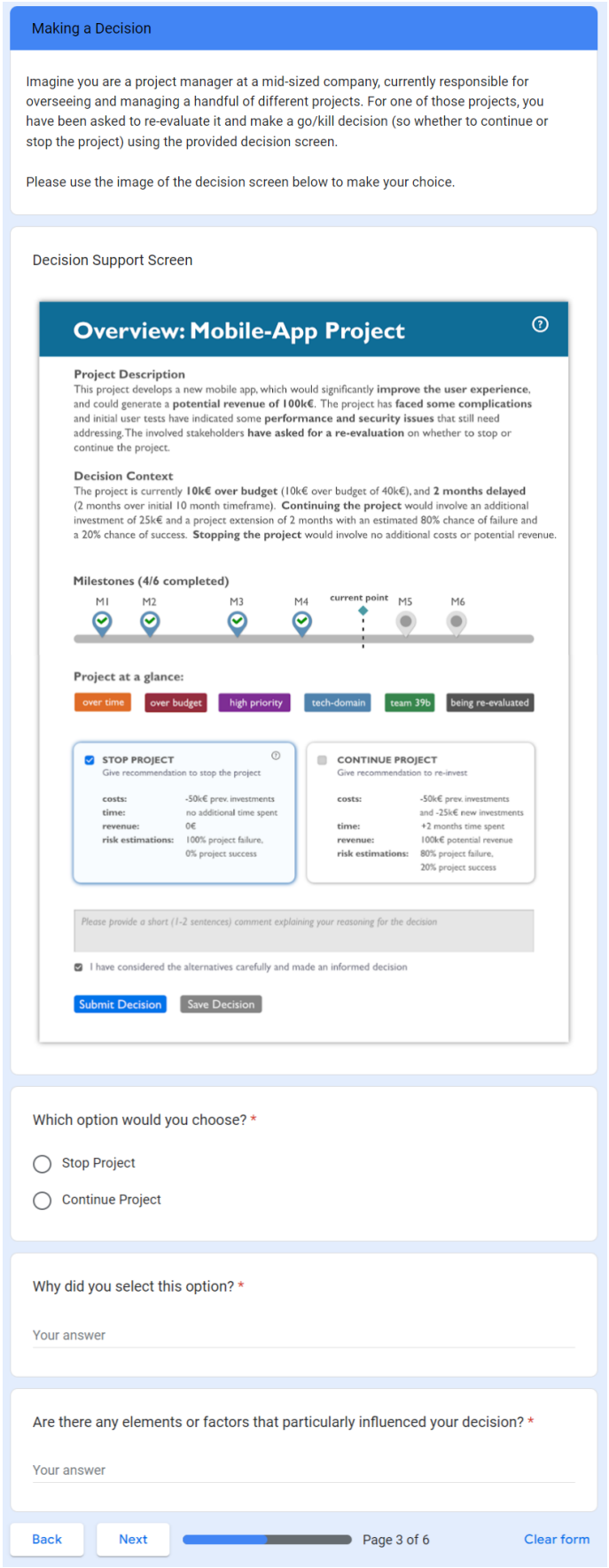

FIGURE 48: Decision Section Questionnaire

FIGURE 49: DSS Usability Questionnaire

FIGURE 50: User Acceptance Questionnaire (Covert Nudges)

FIGURE 51: User Acceptance Questionnaire (Transparent Nudges)

Structure Online Questionnaire User Acceptance

Questionnaire Structure

Category	Step	Description	Data	Duration
Introduction (~5mins)	Formalities	Informed consent section		<2mins
	Demographics	Collect demographic data	- What is your age? - What is your current occupation? - How would you rate your experience with project management?	
	Provide Decision Scenario	Provide the participant with a detailed scenario on why decision is being made		
Making a decision (~5mins)	Decision	Provide the participants with a screenshot of the DSS screen with covert nudges and ask them to make a choice	The selected option	2-3mins
	Follow up on reasoning	Ask participants to elaborate on why they chose which option	- Why did you select this option? - What elements/factors that influenced your decision?	
	Survey on Usefulness	get opinions on usefulness of the used DSS screen	(some questions adapted from TAM: 7pt Likert scale -> strongly agree to strongly disagree) - I think that using such a system would make it easier to make go/kill decisions - I would find this system useful when making go/kill decisions - I think this system would hinder me while	

			making go/kill decisions. - I think the provided information is relevant for making go/kill decisions - I think the provided information is insufficient for making go/kill decisions - Elaborate in open text option	
User Acceptance of Covert Nudges	Highlight Nudges	Highlight the used nudges and explain intent behind them		2-3 mins
	User Acceptance	Survey on perceived manipulation and acceptance of the covert nudges	adapted from Scale of Trust in Automated Systems (Jian et al.); [7pt Likert scale -> strongly agree to strongly disagree] - I think this system is deceptive/manipulative - I am suspicious of this systems intents, actions, or outputs - I can trust this system - I think this system has my best interest in mind - I think the applied strategies are acceptable in this specific use-case • An (optional) open-text option on deception/trust and their opinion on it	
User Acceptance of Transparent Nudges	Highlight Transparency in Nudges	Highlight how the used covert nudges can be made transparent by adding infoboxes and explanations		2 mins
	User Acceptance	Survey on perceived manipulation and acceptance of transparent nudges	Repeat previous section	

FIGURE 52: Overview Questionnaire Structure

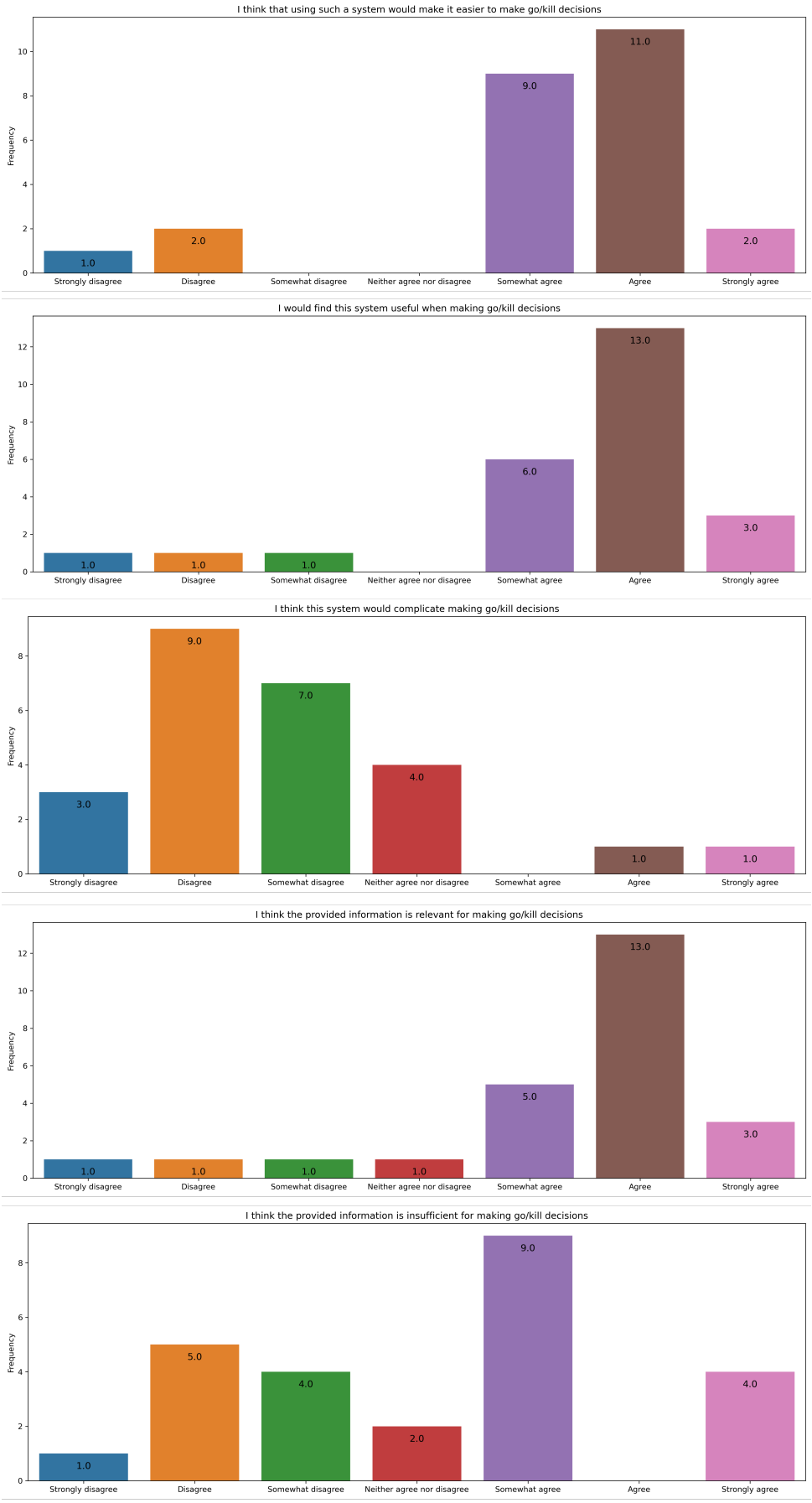

FIGURE 53: Overview of Usability Results

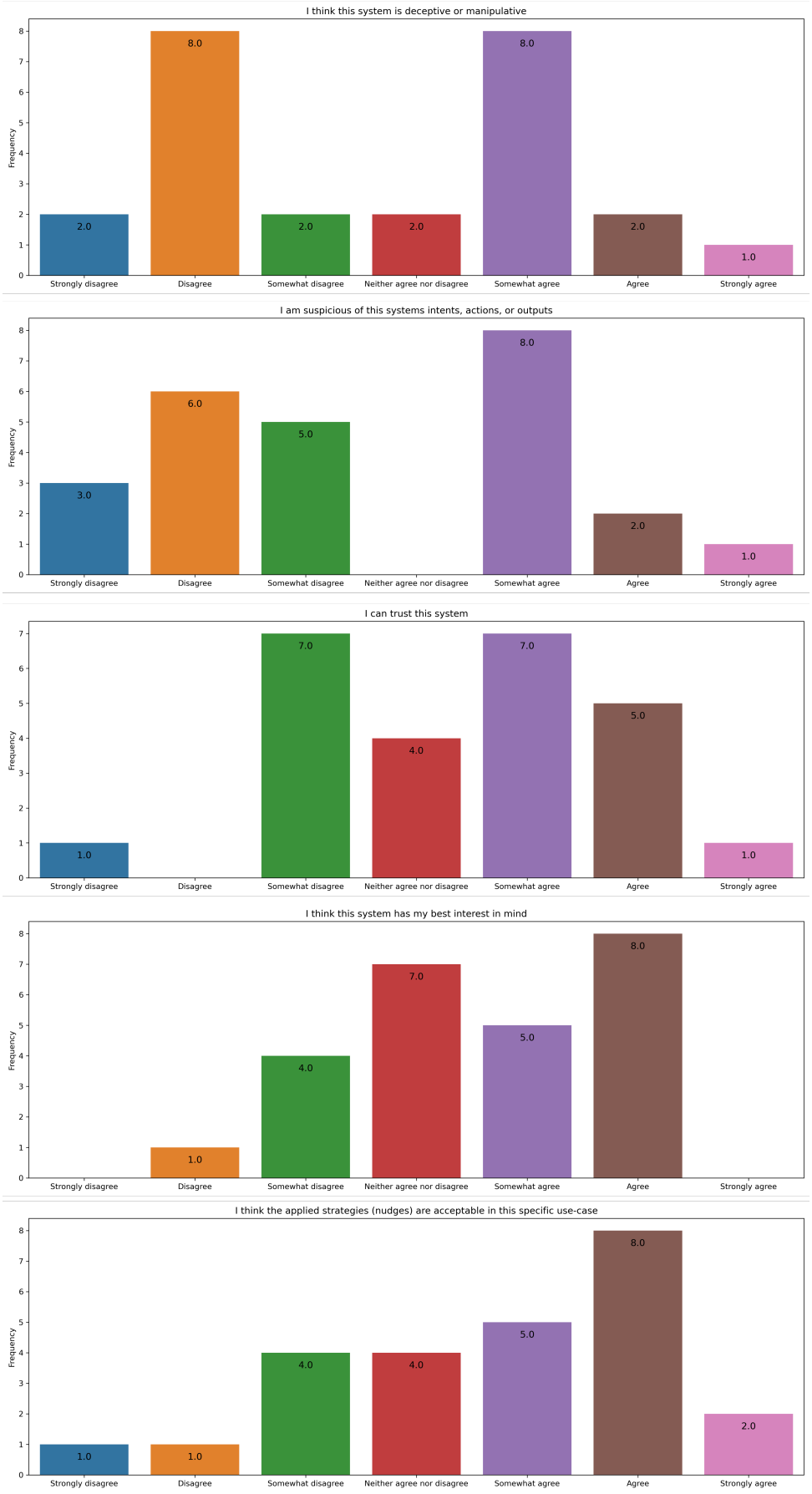

FIGURE 54: Overview of User Acceptance (Covert Nudges)

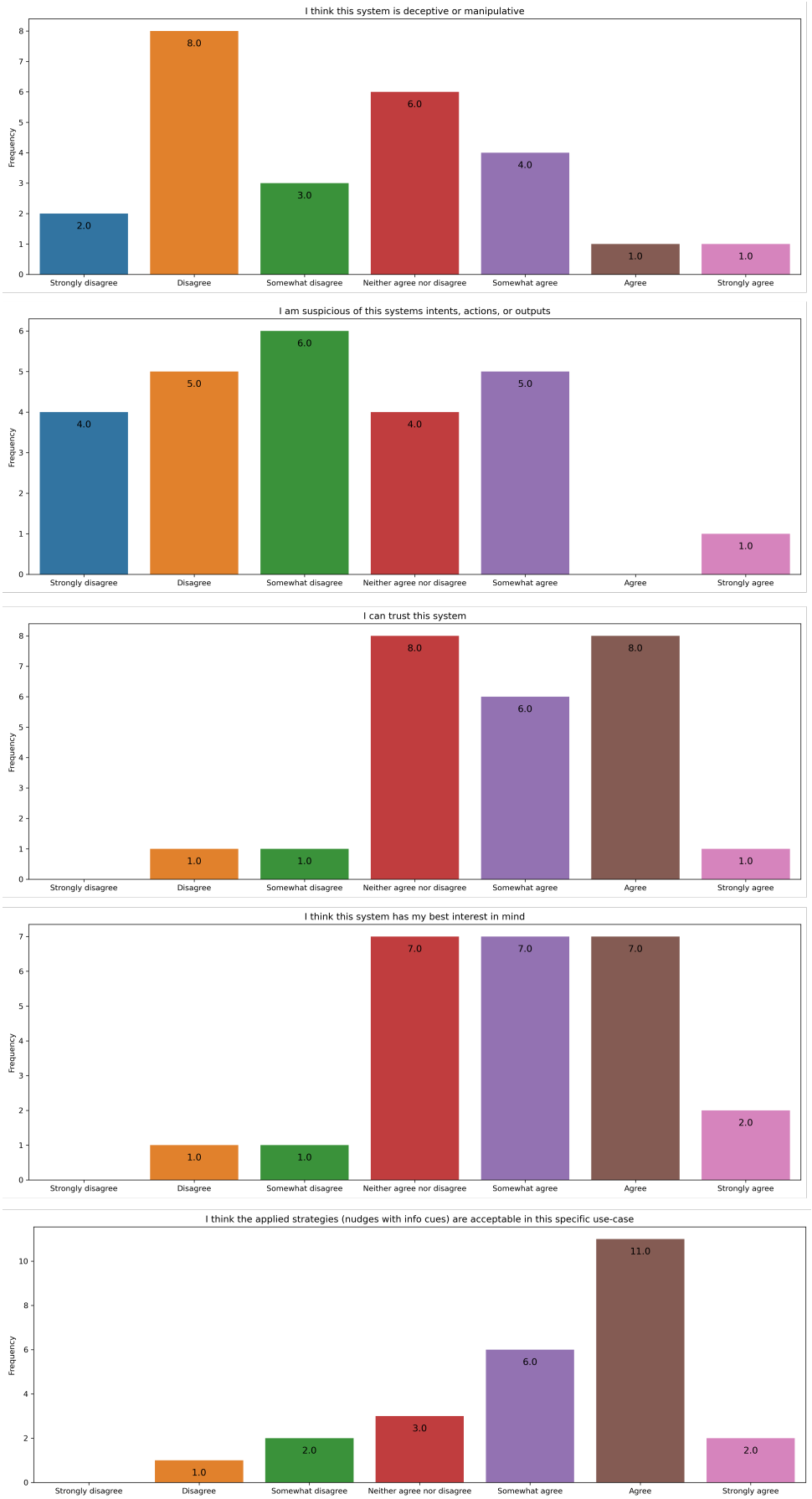

FIGURE 55: Overview of User Acceptance (Transparent Nudges)

Statement	Experience Level	Mean	SD	Mann-Whitney U	p-value
Using system makes go/kill decisions easier	Low	5.44	0.73	63.50	0.6270
	High	5.06	1.77		
Finding system useful for go/kill decisions	Low	5.78	0.67	78.50	0.7117
	High	5.19	1.72		
System would complicate go/kill decisions	Low	2.22	0.67	45.00	0.1192
	High	3.19	1.64		
Information relevant for go/kill decisions	Low	5.89	0.60	89.00	0.3112
	High	5.06	1.73		
Information insufficient for go/kill decisions	Low	3.67	1.94	55.50	0.3497
	High	4.44	1.71		

TABLE 1: Result Comparison Usefulness per Experience Level

Statement	Experience Level	Mean	SD	Mann-Whitney U	p-value
System is deceptive or manipulative	Low	3.67	1.73	73.50	0.9533
	High	3.62	1.82		
Suspicion of system's intents	Low	3.44	1.88	68.50	0.8613
	High	3.62	1.78		
Trust in system	Low	4.33	1.22	67.00	0.7933
	High	4.44	1.55		
System has my best interest in mind	Low	4.22	1.39	51.50	0.2415
	High	4.81	1.11		
Applied strategies are acceptable	Low	4.78	1.39	70.50	0.9537
	High	4.69	1.70		

TABLE 2: Result Comparison Covert Nudges Experience Level

Statement	Experience Level	Mean	SD	Mann-Whitney U	p-value
System is deceptive or manipulative (Transparent Design)	Low	3.11	1.54	61.50	0.5610
	High	3.50	1.63		
Suspicion of system's intents (Transparent Design)	Low	2.89	1.54	60.50	0.5258
	High	3.38	1.63		
Trust in system (Transparent Design)	Low	4.78	0.83	62.50	0.5957
	High	4.94	1.34		
System has my best interest in mind (Transparent Design)	Low	5.11	0.78	79.00	0.7035
	High	4.88	1.41		
Applied strategies are acceptable (Transparent Design)	Low	5.22	1.39	75.50	0.8579
	High	5.19	1.22		

TABLE 3: Result Comparison Transparent Nudges per Experience Level

Statement	Decision Type	Mean	SD	Mann-Whitney U	p-value
Using system makes go/kill decisions easier	Stop Project	5.15	1.63	55.00	0.7429
	Continue Project	5.40	0.55		
Finding system useful for go/kill decisions	Stop Project	5.30	1.56	46.00	0.7959
	Continue Project	5.80	0.84		
System would complicate go/kill decisions	Stop Project	2.95	1.47	57.50	0.6213
	Continue Project	2.40	1.34		
Information relevant for go/kill decisions	Stop Project	5.30	1.59	54.00	0.7966
	Continue Project	5.60	0.89		
Information insufficient for go/kill decisions	Stop Project	4.15	1.76	49.50	1.0000
	Continue Project	4.20	2.17		

TABLE 4: Result Comparison Usefulness per Choice (Stop/Continue)

Statement	Decision Type	Mean	SD	Mann-Whitney U	p-value
System is deceptive or manipulative	Stop Project	3.60	1.70	46.50	0.8330
	Continue Project	3.80	2.17		
Suspicion of system's intents	Stop Project	3.45	1.67	41.00	0.5524
	Continue Project	4.00	2.35		
Trust in system	Stop Project	4.40	1.31	53.00	0.8613
	Continue Project	4.40	1.95		
System has my best interest in mind	Stop Project	4.60	1.19	50.00	1.0000
	Continue Project	4.60	1.52		
Applied strategies are acceptable	Stop Project	4.65	1.60	45.00	0.7540
	Continue Project	5.00	1.58		

TABLE 5: Result Comparison Covert Nudges per Choice (Stop/Continue)

Statement	Decision Type	Mean	SD	Mann-Whitney U	p-value
System is deceptive or manipulative (Transparent Design)	Stop Project	3.65	1.57	77.50	0.0596
	Continue Project	2.20	1.10		
Suspicion of system's intents (Transparent Design)	Stop Project	3.50	1.61	79.00	0.0485
	Continue Project	2.00	0.71		
Trust in system (Transparent Design)	Stop Project	4.75	1.21	33.50	0.2576
	Continue Project	5.40	0.89		
System has my best interest in mind (Transparent Design)	Stop Project	4.85	1.23	38.50	0.4396
	Continue Project	5.40	1.14		
Applied strategies are acceptable (Transparent Design)	Stop Project	5.15	1.23	42.50	0.6161
	Continue Project	5.40	1.52		

TABLE 6: Result Comparison Transparent Nudges per Choice (Stop/Continue)