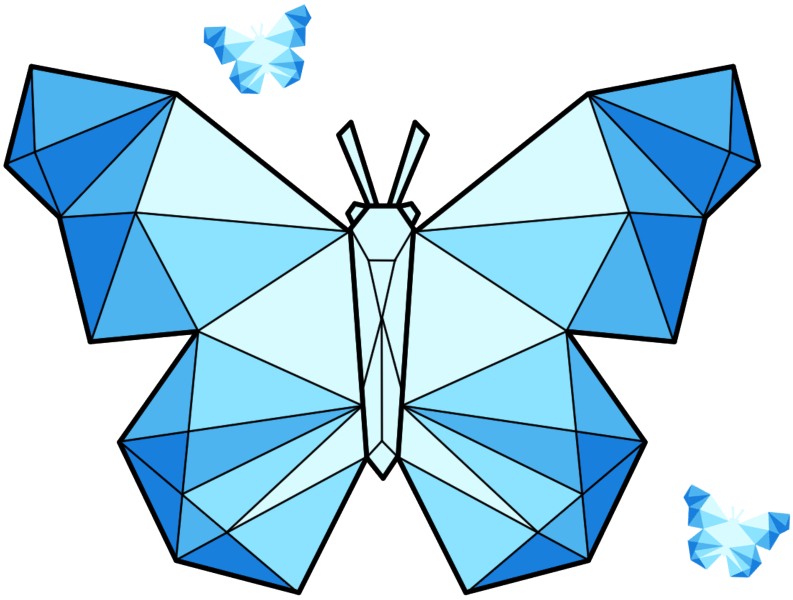
CHALMERS UNIVERSITY OF TECHNOLOGY | UNIVERSITY OF GOTHENBURG



FlyFit: Exploring Abstractness in Visualisations for Fitness tracking data

Evaluating visualisations with different levels of abstractness to support reflection and engagement

Master's thesis in Interaction Design and Technologies

JENNIFER KROGH

Department of Computer Science and Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
UNIVERSITY OF GOTHENBURG
Gothenburg, Sweden 2023

Master's thesis 2023

FlyFit: Exploring Abstractness in Visualisations for Fitness tracking data

Evaluating visualisations with different levels of abstractness to support reflection and engagement

JENNIFER KROGH

UNIVERSITY OF GOTHENBURG

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Computer Science and Engineering
Chalmers University of Technology
University of Gothenburg
Gothenburg, Sweden 2023

FlyFit: Exploring Abstractness in Visualisations for Fitness tracking data
Evaluating visualisations with different levels of abstractness to support reflection and engagement
JENNIFER KROGH

Supervisor: Paweł W. Woźniak, Department of Computer Science and Engineering
Examiner: Staffan Björk, Department of Computer Science and Engineering

Master's Thesis 2023
Department of Computer Science and Engineering
Chalmers University of Technology and University of Gothenburg
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Typeset in LaTeX
Gothenburg, Sweden 2023

FlyFit: Exploring Abstractness in Visualisations for Fitness tracking data
Evaluating visualisations with different levels of abstractness to support reflection and engagement
JENNIFER KROGH
Department of Computer Science and Engineering
Chalmers University of Technology and University of Gothenburg

Abstract

Visualisations are frequently used within personal informatics systems, including fitness tracking tools. The purpose of tracking tools is to help users change their behaviour, but they are often unable to engage and support users. This often leads to abandonment of the tool and the behaviour remains unchanged. Not properly maintaining all elements of well-being, where physical well-being is included, results in a decline in one's quality of life which is why it is important that tools are helpful to users.

This thesis explores visualisations with three different levels of abstraction with reflection and engagement as the focus. The three levels include a classic bar chart as well as two abstract visualisations, designed during this project. A survey (n = 222) was used to evaluate the visualisations. A mobile application called FlyFit with the visualisations included was then developed and used in follow-up interviews (n = 9) to gain a deeper understanding and explore the intended context.

I found that low levels of abstraction offered a better potential for reflection while no significant difference was found regarding engagement. My contribution is insights into the use of abstraction in personal informatics together with suggestions for future work within the field, such as exploring the use of multiple visualisations together, customisation, and different shapes of abstract representations.

Keywords: Interaction Design, Personal Informatics, Fitness Tracking, Visualisations, Abstractness, Mobile Application.

Acknowledgements

I would like to thank the people that, in any way, aided and supported the work of this thesis. First, my supervisor Paweł W. Woźniak for guiding me into this area of research, providing me with useful references, advising me on what to focus on and how to design studies well, lending me the tools necessary to run the studies, helping me structure the thesis, and everything else.

Next, my family and friends for participating in the two parts of the on-the-spot pre-study, providing me with valuable feedback during the design process, Jasmin Niess for giving me feedback on the survey, and Marit Bentvelzen for helping me setting up the first study,

Lastly, all the people who participated in study I and II, making it possible to explore the research question and gain valuable insights.

Jennifer Krogh, Gothenburg, June 2023

Contents

List of Figures

List of Tables

1

Introduction

Self-tracking, or personal informatics, is now part of our everyday life [73]. A large variety of areas are being tracked with trackers having different goals, motivations, and intensity of reviewing their tracking [33, 73]. The most commonly tracked area is physical activity [31], as a way to obtain a healthier lifestyle and change behaviours. However, trackers face many everyday challenges with tracking such as abandonment, lapsing, integrating tracking with everyday life, and making sense of data [31]. One way to assist the user in making sense of data is by using visualisations.

Vision exceeds all the other senses when it comes to humans acquiring information [100]. When presented well, visualisations aid the understanding of data to a great extent. To fully utilize this, the most appropriate visualisation should be used to represent the data. This is currently not always the case, and individual areas should be studied to determine the most useful data visualisation to support users and evoke desirable emotional responses. Previous work has shown that how data is visualised indeed has an impact on this. For example, when visualising unmet fitness tracking goals, rumination is more triggered in multicoloured charts while reflection more often occurs as a result of bar graphs [74].

Reflection is highly important within personal informatics. This is because behaviour change is the most common reason for starting to self-track [33] and because reflection is a key element in initiating the slow change that is necessary for a behaviour to change [91]. This makes it even more important to design personal informatics systems that support reflection rather than undesirable emotions such as frustration and confusion, yet many tools today seem to bring out the latter [64]. Not only that, abandonment of tracking tools is a widespread issue because of their inability to support and engage users, as well as numerous other design issues [23, 40, 43, 73].

Many studies have attempted to tackle the issues with current tracking tools. For example, by comparing several different visualisation techniques [30, 96], by including contextual information about factors that affect physical activity [59], by exploring abstract visualisation [21, 34], and by nudging people towards a specific behaviour on an unconscious level [85]. Although many valuable insights have been discovered through the studies conducted within personal informatics and visualisations, there are still research gaps in the field that need to be addressed to better understand the users and their needs.

In this thesis, the focus is to study the use of visualisations within personal informatics. A concrete visualisation (such as bar charts, line charts, and other standard visualisations typically found in tracking tools) will be compared against abstract visualisations of the same data. These visualisations will be designed and evaluated through two studies, the first being an online survey. Based on the results, interviews will be held which will focus on the use of the visualisations in the intended context. A mobile application will be developed as a means to explore the visualisations during the interviews. From the two studies, I determine that abstractness has an effect on reflection and that different abstraction levels are perceived differently in terms of interpretation, motivation, and appeal.

1.1 Research Problem

Self-tracking generates personal data. Presenting personal data in large quantities in a way that is easily understood by consumers is challenging, as shown by previous research [10]. Because personal data are used in a personal context, the unique requirements and characteristics must be considered when designing for it [48]. There is no consensus within the field of personal informatics on how visualisations for self-tracked data should be designed.

Using visualisations unfit for a specific context is unsupportive and instead frustrates the user. Most visualisations are ineffective for a specific type of data or purpose, which makes validating a design's effectiveness both difficult and necessary [72]. Additionally, humans can be tricked into seeing patterns not actually present in a visualisation, and visualisations must thus be designed with care [95].

Whether abstract or concrete representations are the best option is unclear. An apparent divide can be seen from previous research, where both options have advantages and disadvantages. On one hand, abstract representations can increase privacy, possibly increase the opportunities for reflection, be pleasant to look at, and in some cases create an emotional response. On the other hand, they can also be confusing, create a disconnection with the user, and decrease the possibility for users to spot trends. Concrete representations can help users find specific information and trends, aid the identification of aspects that affects the user's behaviour, they are flexible and people are often comfortable understanding them. However, the low levels of graphical literacy frequent within the general population affect how well graphs can be understood by the average users. They can also be perceived as boring and lack the ability to motivate users. In addition, with current tools (where concrete representations are frequently used) unable to engage users in the long run, the need for visualisation techniques to be reviewed becomes apparent.

Furthermore, abstractness can be viewed on a spectrum and appear in different manners. Previous work has used different levels of abstractness [12] as well as different kinds of abstractness [46]. Since it is clear that concrete and abstract representations come with both advantages and disadvantages, a visualisation that has aspects of both techniques might prove to be valuable.

1.1.1 Research Question

As current personal informatics tools have design issues that lead to user disengagement, and because designing effective visualisations fitting for a specific purpose is a complex task, the use of different visualisations types for self-tracking data will be studied further. With that said, the research question that will be explored throughout the work of this thesis is:

How does abstractness affect self-reflection and engagement in visualisations of physical activity data within the context of personal informatics?

1.2 Aim

This thesis aims to explore visualisations in the context of personal informatics. This will be done by focusing on different levels of abstractness in visualisations of physical activity data. The abstraction levels will be compared against each other and evaluated based on factors important for personal informatics tools and their users. A simple mobile application will also be designed and implemented to use for evaluating the visualisations in a situation resembling the intended context.

The research conducted at the start of the thesis work will be used as a base for the design and implementation parts, where the resulting prototypes ultimately will be used as tools for the evaluation that aims to tackle the research question. The goal is to provide more insight into what can be considered appropriate choices when using visualisations within personal informatics. This could also potentially inspire similar work to be done in other areas, as the importance of how different visuals affect users becomes apparent.

1.3 Stakeholders

This project's primary stakeholders are presented below. All of the user groups presented should be considered in the context of tracking applications as well as the use and creation of those.

Users: Users, also known in this context as trackers, are the ones using the applications where visualisations of personal data are presented. This group would benefit from the research conducted in this thesis since the abandonment of tracking applications are so widespread due to design issues, and this thesis aims to study that aspect further.

Product Owners: Quitting to track or switching between tracking tools are frequent occurrences for self-trackers [33], and a common reason for switching tools is that the users found another they believed was "better". This could be in terms of features, ease of use, accuracy, or something else. Comparing different tools also occurs with trackers who wish to change their behaviour. By creating an application that supports users well, it might be possible to keep users who otherwise would have switched to another tool.

Designers: User research, which is to understand and study people's activities, goals, and the contexts where technology might be useful, usually occurs in the design process of a UX designer [13]. The more research that is contributed to the area of user experience and interaction design, the easier it might be for the designer to discard certain ideas at an early stage if previous research has already shown that there is a better alternative.

Personal Informatics field: The field of personal informatics naturally benefits from research done on important subjects within the field. The aim of this thesis is to contribute with more knowledge on different types of visualisations used for fitness tracking data, which will hopefully be valuable for future research.

Chalmers University of Technology: An institution offering education and conducting research with sustainability in mind. The university will be responsible for publishing the finished thesis and will be acknowledged for the academic outcome.

1.4 Project Plan

The plan for the project was to have the activities divided into five phases. The first four phases follow the process that will be presented in Chapter 3, *Method*, with an additional phase to complete mandatory activities. The five phases are; Background, Design, Implementation, Evaluation, and Finalize. Each phase was planned to have around four weeks dedicated to it. Thesis writing was the only planned activity to expand across all phases.

Background: Phase 1 focused on doing the groundwork necessary to be able to conduct the practical work of the thesis in a proper and educated manner. The activities for the first phase were:

- *Create work plan*: Plan how the work of the thesis should be carried out during the course of 20 weeks. Setting some main deadlines as a way to make sure the work was proceeding as planned.

- *Conduct literature studies*: Review relevant literature, including the research area, related work, and concepts that could be helpful for the project. Studying the literature helped guide the work and provided valuable knowledge that could be utilized in upcoming phases.

- *Review tools for implementation*: Since the implementation phase involves implementing the prototype using a programming language and a framework, it was important to review documentations of those tools to facilitate the process. This activity included reviewing guidelines and components that support best practices.

- *Work on planning report*: Since the planning report had to be done around four weeks after the start of the thesis, it had to be considered during this phase as well. Incidentally, the other activities in this phase contributed to the planning report.

Design: The focus of Phase 2 was to develop the abstract visualisations and design the application. During this phase, the activities were:

- *Ideate*: Based on the knowledge acquired during the literature studies, several design ideas were generated. Methods, such as brainwriting, were used during this activity.
- *Sketching*: Both to create design alternative designs and the first low-fidelity prototype, sketching was used. The sketching was partly combined with ideation since new ideas were developed during the process.
- *Prototyping*: This activity refers to the final prototype of higher fidelity that was used as a guide when implementing the application. It was created as a digital prototype using Figma, and colours, placements, and text were more refined than previous prototypes.

Implementation: Phase 3 focused on the software development aspect of the thesis. The activities for the third phase were:

- *Write code*: Implementing the application required writing the code behind it. This was done using Dart as the programming language and Flutter as the framework.
- *Create visuals*: Alongside writing the code, the visuals used to visualise the data were created. This was done in Clip Studio Paint.
- *Test code*: Before the application could be used in a study, the code had to be tested. Errors in the code could affect the upcoming evaluation and had to be avoided.

Evaluation: The focus of Phase 4 was to evaluate the designed visualisations as well as the implemented application featuring them. During this phase, the activities were:

- *Prepare studies*: Before the studies could be conducted, some materials needed to be prepared such as a survey, interview questions, etc.
- *Run studies*: The studies were run separately and the results from the first were used to guide the second study.
- *Analyse and report results*: The last part of the evaluation phase was to analyse and report the results. This was done as soon as enough participants had participated in the studies.

Finalize: Phase 5 focused on finalising the thesis and project work, as well as conducting the other mandatory activities. The activities for the fifth and final phase were:

- *Participate in 2 presentations*: As per the requirements, time was dedicated to participating in two presentations. No preparations were necessary for this.

- *Opposition*: Reading the other's report and writing feedback was required for the opposition to be completed.
- *Prepare and conduct presentation*: A presentation of the thesis project was prepared, practised, and presented. The feedback received was then integrated into the report afterwards.
- *Thesis writing*: Writing the report took place during the whole thesis work. This was of course because writing soon after the activities had been conducted was more effortless as the activity and outcome were more clear in one's memory. In addition, it allowed for more refinement as the oldest parts could be rewritten several times over.

2

Related Work

This chapter introduces relevant related work from the field. The information will initially be about general theories such as interaction design and design principles. After that, the research area will be presented: personal informatics with a focus on goals and reflection, visualisations with a deeper look at the personal context and different visualisation options, and emotional design. Lastly, the implementation domain and some challenges with it will be highlighted.

2.1 Interaction Design

A way to describe interaction design is to describe it as a dialogue between a product and a person [55]. Since the dialogue occurs over time, it is difficult to structure and it requires an understanding of human behaviour. Another way to describe the concept is that it is the process of creating and deciding use-oriented qualities such as aesthetic, structural, and functional aspects of a digital product [64]. It is about designing interactive products, and these products should support people's everyday communication and interaction [81].

Despite the purpose of many products being for users to interact with them, many everyday artifacts don't appear to have been designed with the user in mind [81]. Even though interaction design principles have been around for decades, the same errors occur repeatedly [25]. Products that should be effortless and enjoyable instead create frustration and annoyance, as if there are limitations or a lack of knowledge within the field [64].

Indeed, interaction design is a difficult and complex task, and one challenging aspect is that digital artefacts are concerned [64]. Since digital products are constantly evolving, there is hardly any time for reflection. Just take the revolutionary device released in 2007 as an example, the iPhone. In the 16 years since its initial release, 16 generations, and 38 models have been produced [49].

Usability is often used to evaluate interaction design [55]. To break it down even further, usability can be measured through six usability goals; Effectiveness, Efficiency, Safety, Utility, Learnability, and Memorability [81]. Others suggest that usability should cover effectiveness, efficiency, and satisfaction [14]. Evaluating the usability of a product can be done during many stages of the design process, and by doing so, design problems can be discovered early.

As previously mentioned, efficiency is often used to evaluate artefacts within interaction design. However, as technology evolves, new ways of using it emerge. For example, reflection is an important element of behaviour change (two aspects that will be described further below), and these are things that take time. Current interaction design practices are not fit for products with these kinds of aims, and a design agenda known as *Slow Technology* has emerged as a result of it [45].

By developing slow technology, where the focus should be on the slowness of appearance and presence as well as aesthetics of materials and simplicity [45], *Slow change* can occur [91]. Changing is difficult, and the designer shares the responsibility of the change process with the user. Designers framing change as a simple process risks feeding users with unrealistic expectations, which often results in users giving up. Some design themes to think about when designing systems to nudge people in taking the first step towards change are; designing a system that can change as users, motivations, and needs inevitably change, using realistic and achievable goals, accommodate for slip-ups that often happen, and finding ways to motivate consistency.

2.1.1 Design Principles

Several design principles within interaction design should be used as guidelines when designing a product. These principles encourage designs that create positive experiences and support users' goals and needs. In addition, they help translate requirements into interface behaviours and structures during the design process. Cooper [25] divides the principles into three categories; Conceptual, Interface-level, and Behavioural Principles.

- *Conceptual Principles*: These principles help define the way digital products should be and their structural fit in the context of user requirements. Some examples are reducing work, digital etiquette, platform, flow, and eliminating excise.
- *Interface-level Principles*: Effective strategies for navigation, organization, and communication of information and behaviour are described by these principles. Dialogues and controls are some examples, but there are also different principles when designing for web, mobile, desktop, and other devices.
- *Behavioural Principles*: Principles describing the behaviour of a product, both in specific and general contexts, fit into this category. Examples of this are informing decisions, preventing errors, and visual design.

Sharp et al. [81] further describe design principles as having evolved from experience, common sense, and theory-based knowledge. The authors list a few of the most common principles; Visibility (functions should be visible), Feedback (let users know when an action has been accomplished), Constraints (restricting user interaction at given moments to avoid mistakes), Consistency (similar tasks should have a similar design), and Affordance (objects should give clues how to interact with them). On the other side, applying several design principles can be a challenge as some of them can be in conflict with each other, which will lead to the designer having to make a trade-off.

2.2 Personal Informatics

Following the rise of technology, the frequency with which a user can track his behaviour has reached further than ever seen before. For centuries, self-tracking has been done using only pen and paper [73], but now because of sensors in our mobile phones, the tracking can begin before we even decide we want to start tracking.

Personal informatics can be defined as collecting and reflecting on personal data as a means to understand one's behaviour better [58]. The core aspects of personal informatics are collection and reflection, and by collecting relevant personal data, the user can develop insightful reflections [57]. One personal informatics model frequently used within the field is the five-stage model; Preparation, Collection, Integration, Reflection, and Action [57]. Personal informatics is also commonly referred to as self-tracking.

People often start to self-track to change behaviours by monitoring them [33, 81]. Trackers are encouraged to change their behaviours through graphs displayed based on their data. By self-tracking, trackers can both gain insights into their behaviours and increase their motivation for changing [33]. Although behaviour change is the most common reason for starting to track, curiosity, self-understanding, and simply having a particular behaviour recorded are frequent reasons as well.

Research has shown that many personal informatics tools are unable to engage the users, which could be a factor as to why early abandonment is so widespread [73]. Forgetting to use a tracking tool or because the upkeep is tedious are some reasons for lapsing, and repeated lapsing often results in the user quitting to track [33]. There are also several design issues that hamper long-term engagement. A few examples are that the system goals don't match the user goal, that people achieves or changes their goals over time, and that information is difficult to interpret or access [43]. In addition, to develop effective personal informatics systems, there are a few design considerations to take into account; systems should be viewed as a whole as barriers transfer to upcoming stages, tools should be flexible in terms of data collection, tools should have a mix of user-driven and system-driven approaches to be effective and enjoyable, and multifaceted systems might provide valuable insights and should therefor be explored [57].

How users interact with their tracking devices can generally be divided into glances and engage sessions. Glances are quick looks at the data while engage sessions are longer where the users reflect on the feedback. A 10-month study of an activity tracker showed that the majority of the interactions were glance, which also increased significantly over time [40]. Engage sessions were most common in times when goals had not been achieved, and over time these sessions became less frequent. This study also showed, as previously mentioned, the issue with long-term engagement. Only 14 % of the participants used the application for longer than two weeks, and by the end of the study, all participants had stopped using it.

Even though there exist many tracking tools already, it is obvious that they are not optimal or even desirable to users when it comes to design and functionality. This is shown by how many dedicated self-trackers, also known as Q-selfers (quantified selfers, the term used for self-trackers within the quantified self movement), have built their own tools and designed unique visualisations [20]. Some reasons for doing so are; existing tools do not support their needs, to centralize their tracking otherwise spread out on different applications, and because most commercial tools do not support both tracking and exploring data with a single tool.

2.2.1 Reflection for Changing Behaviours

One aspect of designing personal informatics systems is to decide whether to make the system more reflective or persuasive [70]. Reflective systems help users with reflection and self-regulation in their journey towards achieving self-chosen goals, while persuasive systems push people to behave in a certain way, using prescribed goals to achieve a quick change [80]. Designing a system that supports reflection should display data in a neutral way.

Reflection has an important role in the process of changing behaviours [58]. By gaining valuable insights, people can reconsider their choices which opens the way to possibly changing a particular behaviour [80]. Within personal informatics, users often reflect on their health status and lifestyle choices [2]. The ability to reflect on one's data and the value of it is affected by what data is collected [58]. Many trackers believe that in order to fully reflect on one's performance, more than one type of data must be tracked despite having one primary data type as the main interest [2].

Depending on in which phase of self-tracking the user is currently in, which type of feedback is provided by the tool, and at what time the feedback is received, different types of reflections can be done [19]. Reflection-in-action occurs when the activity is being performed, and reflection-on-action occurs after an activity has been performed [80]. While reflection-in-action is done through real-time feedback, allowing the reflection to happen at the right time, reflection-on-action is done through aggregated feedback and can thus provide connections and long-term patterns in addition to the raw information. A challenge with reflection-in-action is to design tools in a way that supports users without interrupting, and a challenge with reflection-on-action is how to represent data from a particular activity. Typical reasons for wanting to reflect-in-action is to be able to take immediate actions towards reaching one's goal, for example, change route, adjust the activity pace, or switch activity [2].

2.2.2 Goals in Personal Informatics

When trackers reflect on their data, the insights people might be interested in can be divided into six categories; Single Value, Multiple Values, Goal-Based, Comparison (Single), Comparison (Multiple), and Motivational [2]. Another study also mentions goals as one of the questions people usually have about their personal information [58]. Different types of data are needed for different purposes, questions, and insights, which in turn affects how a tracking tool should be designed to support these needs.

Goal-driven motivations are one of the main reasons for tracking fitness data [2], and people having a goal in mind is also often the motivation for starting to track other areas as well [33]. Over time, goals evolve with the user [58, 75] which requires tools to be designed so they can support long-term usage as well as transitioning between goals [33].

Specific goals motivate people to do better than when presented with do-your-best goals, as the latter is ambiguous and allows the acceptable performance level to be of a wider range [62]. The motivation to act is a combination of anticipated satisfaction, believing rewards will be the result of performance, and believing effort will give rise to the necessary performance that will lead to obtaining the rewards. Goals affect performance, and the goal-performance relationship becomes the strongest following commitment. For people to commit to a goal, the outcome of the goal must be of importance, thus achieving the goal becomes meaningful, and people must believe that it is possible to achieve the goal.

Summary feedback, revealing the progress towards the goal, is necessary for goals to be effective [62]. Not knowing the current progress makes adjusting the level of effort difficult. In addition, people tend to increase their effort when finding out they are below the target. It is also common for physical activity trackers to check their data immediately following the collection of it [57], for awareness, to know their current status, and for correcting their behaviour towards reaching the goal [58, 86].

The process of setting goals usually includes a self-assessment and making sure the goal is Specific, Measurable, Achievable, Relevant, and Time-bound (SMART). It is also important to measure the process toward the goal regularly. Several methods for goal-setting can be used, including self-set, assigned (by for example a fitness expert, medical expert, or national recommendations), participatory (working with someone to set a goal), or guided (receiving multiple options to choose from), where the last two could also be done with a fitness expert or a medical expert [22]. A study regarding goal-setting showed that participants preferred self-set goals and working with a fitness expert, through participatory or guided goal-setting. Assigned goals were not appealing to most participants as those do not take objectives, constraints, and individual abilities into consideration.

2.3 Visualisation

Techniques to communicate a message through diagrams, images, or animations are known as visualisations. Graphs, maps, road signs, and tables of figures are just a few examples, and they are used to assist humans in making sense of stuff. Humans have limitations in their internal cognition and memory [72]. By using visualisations, so-called external representations, the perceptual system can be utilized while the usage of internal cognition and memory can be offloaded. To be useful, visualisations must be presented in a way so that people who use them can understand them [81]. It is necessary to understand the context and how something is represented to draw meaning from data.

A powerful way to facilitate self-reflection and to support users in revealing meaningful insights is data exploration through visualisations [19]. But before any insight can be derived from data, it must be understandable, interpretable, and accessible [48]. Already at small scales, data representations can be misleading. For example, truncated scales on axes can be enough to confuse the person trying to interpret the data [81]. Other common and misleading visualisation practices within the scientific world are axes starting above zero, unnecessary 3D usage, and a lack of uncertainty information [95]. Consider the bars presented in Figure 2.1. One tool appears to be twice as good as the other, while in reality, the difference is only a few per cent.

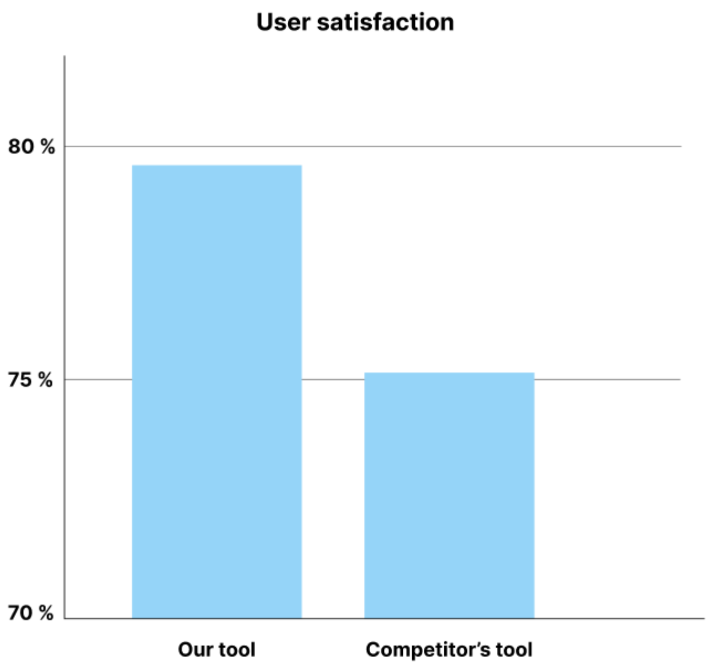


Figure 2.1: Here I have created a bar chart showing how visualisations can be misleading.

The result of a number of design workshop sessions, where professional designers were asked to design glanceable interface visualisations for small screens to aid in the understanding of physical activity data, showed that the designers preferred clean designs with minimal text use [2]. This was achieved by using icons instead of labels, using maximal space available, and using familiar visualisations such as bar and doughnut charts.

2.3.1 Personal Context

Data generated from self-tracking is called personal data. It is constantly changing as new data is generated, and the way people engage with it has yet to be fully explored [65]. Aspects such as how the data becomes meaningful, how it loses meaning, and how it can provide reassurance, comfort, disappointment, or frustration all add to the complexity.

Many publications within personal informatics have discussed the challenges in making sense of personal data [31], and one of those challenges is the personal context [48]. Because of the personal context, new challenges appear as compared to when visualisations are used in a professional context. Thus, there is a need for visualisation and interaction techniques to be reviewed in order to have success in the personal context. One challenge is the inexperience and limited literacy people may have with visualisations, compared to professionals who usually are well accustomed to the use. To accommodate this, there have been instances where different visualisations have been used to represent the same data [81], depending on the intended consumer, such as the general public or experts.

Whether a person is a novice or an expert user in regards to visualisations is not the only user profile aspect that influences what kind of representation is most useful. For example, designs should align with a person's previous experience with tracking as well as the perspective the person has on tracking [32]. A user who has tracked for a long time, and thus has more data collected, is likely to prefer visualisations that show their long-term trend [32] and where they can look for patterns over time [58]. On the other hand, short-term trackers often prefer visualisations that show their day-to-day use [32], and in some cases seem to be uninterested in historic data all together [40, 86].

2.3.2 Concrete Visualisations

A study showed that people generally prefer tables and graphs when presented with their activity data [30]. These types of visualisations were most useful when participants wanted to find patterns and identify aspects that influenced their behaviour. Line and bar charts, which are examples of graphs, are common visualisations in many areas [2]. They are often used to represent trends and historical data. Because they are standard representations, many people are comfortable interpreting, reading, and understanding them. These types of visualisations have also frequently been used in studies about personal informatics [59, 74, 96], and when Q-Selfers want to convey data-driven insights during conferences [18, 20].

Other popular options within information graphics are pie and doughnut charts [52, 59]. Despite that, the visualisation community generally recommends against their use [93]. Radial bar charts have been used and compared against bar charts, and the results showed that the former performed worse in terms of supporting reflection [74]. In addition, radial bar charts can be challenging for users to understand because the user has to compare circles with different areas and radiuses, which can be difficult and confusing [93].

The advantages of graphs are that they are useful in helping people understand patterns [23]. Because of the diversity of graphs available, a compact visualisation can be created to communicate extensive information. Graphs are a flexible form of displaying self-tracked feedback where the possibility to make them interactive greatly contributes to this. A disadvantage of using graphs is the low levels of graphical literacy that exist within the general population, with a large part of the population already having difficulties with basic numerical concepts. This leaves questions regarding how effective graphs are for the general population in terms of being the main form of tracking feedback.

2.3.3 Abstract Visualisations

Ambiguous, ambient, and abstract visualisations are used in many tools to encourage reflection, as a way to involve users in a sense-making process that is open-ended [80]. These types of visualisations have also been used in several studies, and one study used flowers as a metaphor [21], where the goal was to use the aesthetic representations of physical activity data to encourage behaviour change. The result showed that the glanceable interface kept users active, while the test group who did not have it significantly dropped their level of physical activity. Another study used informative art [34] in a public setting where the goal was not for the display to change behaviours. Instead, people's experience with it was the main interest. The study showed that charts were more effective for specific information while the participants found abstract visualisations more appealing when simply glancing at the display.

Another study relevant to this thesis is one where the classic bar chart was compared to a Virtual Pet (abstract) and a Circular Ringmap visualisation [96]. The three visualisations were evaluated based on task performance and the experience reported by participants. Results showed no difference between the visualisations in time spent on tasks, however, the Virtual Pet sparked an emotional response while the bar chart was better in terms of readability and awareness. The Circular Ringmap was deemed unhelpful.

Several design strategies for technologies intended to change behaviours have been identified [24], and one of those strategies is to use abstractions. According to the authors, data abstractions are more effective in encouraging users to reflect on their behaviour compared to displaying raw data. By using abstract representations, privacy is increased which allows the user to engage with the representation in various situations, increasing the opportunity for reflection. On the other hand, the majority of participants from one study believed that the excess of abstract representations used in commercial tools are unhelpful [83]. Many of them preferred more concrete representations in which they could recognize themselves.

Creating engaging and motivating abstract visualisations is a challenge in itself. When working with a metaphor representation, proper variations of the visualisation is needed or there is a risk of users getting bored quickly [41]. In addition, depending on the user profile, vague information may not always be the best option. As tracking often is associated with exact numbers, users may expect the feedback to be presented accordingly. Lastly, if there is not a clear connection between the tracked data and the metaphor representation, it might not be as successful and the design decision might appear to be arbitrary.

The main effects evoked from users from the use of ambiguity can be categorized as; curiosity, social engagement, play, immersion, awareness, understanding, reflection, and creativity [46]. Relevant for this thesis is reflection, and in several studies, participants have been reflective following interaction with an ambiguous design [60], including cases where well-being has been the focus [76, 77]. However, there have also been studies where the aim was to evoke reflection using ambiguous designs, but failed to do so and instead evoked confusion and frustration [46]. With that said, to what extent ambiguity and abstract representations can be useful for reflection within personal informatics is still unsure.

2.4 Emotional Design

Designing products that can evoke specific emotional responses, and how to use emotive feedback to change behaviour, are both aspects of the area known as Emotional Design [81]. Emotions are divided into two types; Automatic emotions, which are short-lived and simple, and Conscious emotions, which are long-lived and complex. By understanding the way emotions work, automatic or conscious emotions can be triggered depending on how the user experience is designed.

Three different installations were used in a study to find out if it was possible to change behaviours or encourage specific ones [85]. When choosing between taking the stairs or the elevator, a significant amount of people changed their behaviour following the instalment of the ambient displays. Even though a majority of people reported that they had not changed their behaviour, the logged data showing the actual behaviour suggest that the displays may have affected people at an unconscious level.

2.5 Mobile Applications

Mobile devices, with their sensing capabilities and high-resolution display screens, are powerful tools for personal informatics and for managing health in everyday life [23]. Yet 74 % of health applications are abandoned within 10 uses, where insufficient usability and design are key factors for the abandonment. Evidently, there is a need for design consideration when attempting to design effective mobile wellness systems.

Besides the considerations necessary when designing specifically for a mobile wellness application, some aspects need to be considered when designing for a mobile application in general. Some of those aspects include input method, screen size, and use context [25]. Buttons must have a large enough "hit area" to avoid mistakes, and navigation through content must be contemplated due to the constraints of the screen size [81]. Mobile application development also shares similar issues with the development of embedded applications (software for specific functions permanently inside a device, for example, a microwave). Some of those issues are performance, security, and storage limitations [101].

2.5.1 Material Design

With the purpose of supporting best practices of interface design, Material Design is an adaptable system consisting of components, tools, and guidelines that support the process of interface development [39]. It is built by Google developers and designers, and it includes User Experience guidance for both mobile phone applications and the web.

With Material Design, adaptive, expressive, and personal experiences can quickly be built. In addition, one foundational design value is accessibility by default, as understanding the range of human experiences can reduce costly redesigns and design debt. The accessibility standards included in the components are designed to assist the development of products with inclusive design. Some examples of available components are Buttons, Cards, Dialogs, Lists, and Navigation bars.

3

Method

Löwgren and Stolterman [64] believe that the design process is too diverse and complex to generally describe, and that to be a thoughtful designer, one must remain critical towards design process descriptions. Rather than completely adopting a design process, a designer should appropriate aspects based on an understanding of design and by using reflective thinking.

With that said, the design process for this project will be inspired by the model presented by Sharp et al. [81]. The authors describe the interaction design process in terms of four activities; Discovering Requirements, Designing Alternatives, Prototyping, and Evaluating. Ultimately, this model is based on parts of the Double Diamond process [26].

The reasons for choosing this methodology as inspiration are that it fits fairly well with the product that will be developed during this thesis, it is created for interaction design, and a simple process model is adequate when only a few designers are involved [81]. Other models such as the Double Diamond process [26] and the Design Thinking process [35] were considered but discarded because of some of the mentioned reasons. Nevertheless, the chosen model will not be followed rigorously and adjustments have been made to fit the work of this thesis where the domain is HCI. The design process for this thesis will be divided into four phases; Background, Design, Implementation, and Evaluation.

3.1 Background

Known as "Discovering Requirements" in the model by Sharp et al. [81], this phase will be about understanding the user and how the interactive product could provide support. Proper literature research will be conducted to understand what has previously been done in the field, where the research gaps are, and how to add to the current research. Details about the outcome of this phase can be found in this section and in Chapter 4, *Design & Implementation*, while the information gathered from the literature research is compiled in Chapter 2, *Related Work*.

3.1.1 Human-Computer interaction

The field of human-computer interaction (HCI) is, as the name states, about human interactions with computers. The domain can be defined as the design, implementation, and evaluation of computer systems that are interactive and intended for human use [92]. As an area originally within computer science, the focus lies on the interaction, but it has since expanded and is now a multidisciplinary field including cognitive science and human factors engineering as well [36].

The area of HCI overlaps with interaction design, and several goals and principles describing what good design is can be found in both. Efficiency, effectiveness, enjoyment, and easy to use are core aspects within HCI and the process of designing interactive computer systems [28]. HCI also overlaps with User Experience Design, but the former is more focused on understanding users through empirical development by using scientific research while the latter is more industry-focused [36].

As stated, HCI is research-driven. Work within the field results in one or more of the following contributions; Empirical, Artifact, Methodological, Theoretical, Dataset, Survey, or Opinion contribution [103]. The work from this thesis will result in an Empirical contribution and partly an Artifact contribution.

3.1.2 Literature Review

Reviewing the literature is an essential aspect of academic writing [15]. Collecting previous research on a specific topic, gathering the essence, and drawing connections between the references will help inform the current project. Focusing on the references most relevant to the project should be the guiding factor. In addition, writing a summary of what each work says is not the purpose, but rather focusing on the collected work as a whole [54].

The benefits of literature reviews are many; it gives a general overview of a topic previously unfamiliar to the researcher, reveals what research has already been conducted, provide new ideas that can inspire new research, and enables new research to be placed in a larger context [54].

It is important to be critical regarding the choice of sources, to make sure they are both relevant and credible [15]. Moreover, due to the massive increase in knowledge production, doing literature research in an accurate way has become progressively complex [94]. For that reason, it is suggested to follow guidelines when doing a literature review. Some guidelines are; to summarize each study's main claim in a short text, group studies based on categories, be selective about what studies to include and write about studies with similar claims together rather than in separate paragraphs [54].

Using the Method: In order to collect relevant information from the literature to use in this thesis, a number of different approaches were used. Google Scholar [38] was the main search engine used to find books, papers, and articles within the field. These potential sources were screened based on their relevance to the project as well as the number of times the sources had been cited, to get a sense of their credibility. Chalmers Library [78] was used to get access to the full version of the books that were deemed relevant.

A mapping of over 500 papers on personal informatics has been done and presented in a paper by Epstein et al. [31]. Their labelling of papers is available in a database. Both their paper and the database were used to find relevant papers on personal informatics as well as to get an overview of the current standing within the field. The focus when reviewing the previous personal informatics studies was ones that used mobile devices as the displaying and tracking tool and studies that focused on tracking physical activity. The second most prioritised areas were studies that used websites and/or that explored tracking users' daily life, health, and well-being. Studies exploring abstractness were also deemed highly relevant.

From the papers and books found through Google Scholar and the personal informatics database, additional sources were discovered by reviewing the references used in those. In addition, old master's theses in Interaction Design, from both Chalmers University of Technology and other universities, aided the search for sources. These sources were mainly about methods and to some extent related work. Books introduced in previous courses were also frequently used and the guidelines presented above were taken into consideration when conducting the literature review.

3.2 Design

Several alternatives will be designed and considered in this phase before a decision about which option to proceed with is made. Ideation is therefore a central aspect, and methods for ideation will thus be used. To consider multiple alternatives is a fundamental principle of design [72]. If the consideration space is too small, there is a possibility that only "okay" or poor solutions are considered, and good solutions might be missed. Chapter 4, *Design & Implementation* describes the process for this phase.

3.2.1 Brainwriting

To settle for an initial idea because it is "good enough" is not good enough when striving to achieve interaction design goals. To consider alternatives is crucial during a design process in order to produce the best possible result. Several methods can be helpful during this stage, and brainwriting is one such method. The technique is used to generate, refine, and develop ideas, similar to brainstorming [81].

When using the brainwriting method, ideas are written down instead of spoken out. Some suggest that the method can produce even more ideas than the traditional brainstorming method since more ideas can be generated in the same amount of time due to participants writing simultaneously [102]. Similar suggestions for conducting a successful brainwriting session applies as when brainstorming, such as; not overlooking silly ideas, defining the problem to keep the focus, knowing the user, and using ideas to create new ideas [81].

The advantages of brainwriting are that several ideas can be quickly generated, minimal facilitation and training are needed, and anxiety, conflicts, and conformity can be reduced [102]. On the other hand, handwriting might be difficult to transcribe, some find it hard to express themselves in writing, and it is less social than brainstorming.

Brainwriting is also known as "individual brainstorming" [102] and it is for that reason brainwriting was chosen over brainstorming. While both methods generate similar results, brainwriting is naturally a better fit when there is only one researcher involved, such as in the case of this thesis.

3.2.2 Sketching

Sketching is highly efficient to realize initial ideas. During a design process, sketching can be used as a powerful method for reasoning, thinking, and exploring opportunities [9]. The goal of sketching is to reveal ideas by transforming them into tangible information for others to grasp. According to Buxton [16], what defines something as a sketch can be determined through several attributes. Some of them are; timely, quick, inexpensive, disposable, minimal details, and explore rather than confirm.

People often feel self-conscious about their drawings and thus refrain from participating in sketching [9]. However, it is important to note that a drawing doesn't have to be advanced in any way to be able to communicate ideas visually. In fact, the more ambiguous a sketch is, the more room is left for different interpretations [16]. Developing new and different interpretations is a key purpose of sketching, and sketching in turn is a central part of design learning and thinking.

3.3 Implementation

The third activity is described by Sharp et al. [81] as the prototyping phase. As the look, feel, and behaviour of interactive products are central design aspects of interaction design, users need to interact with the product to be able to properly evaluate it. This phase will be about implementing the application that will be used during the evaluation phase. The process for this phase is described in Chapter 4, *Design & Implementation*.

3.3.1 Prototyping

Prototyping is a qualitative method commonly used during a design process. It is used to develop and evaluate ideas within a design team but also with users [15]. For interface design, interaction techniques, aesthetics, and content can be examined through the use of prototypes [98]. Since a prototype is limited in the questions it can answer, it should be developed with a key issue in mind [81].

Low-fidelity: Prototypes can vary in resolution or "fidelity". How easy it is to distinguish the prototype from the final product is what fidelity is referring to [98], and the fidelity level can be anything from low fidelity to high fidelity, and everything in between [15]. Low-fidelity prototypes are generally used during ideation to test early concepts. This makes it easier to do iterative changes following timely feedback as well as having a concept proposed for review. Paper prototyping is an example of a low-fidelity prototype often used within software design. Some advantages with low fidelity prototypes are that several design concepts can be tested and revisions can be quickly done, while disadvantages are limitations with the flow and navigation, as well as limited usefulness for usability testing [81].

High-fidelity: Being more complete and refined, high-fidelity prototypes closely represent the intended final product [15]. Since more care is put into the aesthetics and because some basic functionality often is in place, high-fidelity prototypes are commonly used during later stages of the design process. Usability and interaction can be evaluated accurately and the prototype can often provide a real user experience. User-driven, fully interactive, and almost complete in functionality are some advantages with high fidelity prototypes, while some disadvantages are them being time-consuming and resource-intensive [81].

3.4 Evaluation

Evaluating prototypes lets the designer determine a product's acceptability and usability. This is often measured in user experience and usability criteria [81]. Evaluation can also be used to test hypotheses, compare design alternatives, and find out more about how people react emotionally to the use of a product. The results can help guide what is needed for future work. For this project, evaluation is the fourth and final phase, divided into two parts; Study I and Study II. Details about the design and purpose of each study can be found in Chapter 5, *Study I: Online Survey* and Chapter 6, *Study II: Interviews*.

3.4.1 Study Design

An aspect of planning a study revolves around choosing between using a within-subject or a between-subject design. This refers to how the test condition are administrated amongst the participants.

Within-subject: If participants are tested on each of the test conditions, the study design is within-subject [66]. This design is also known as repeated measures as the measurements are repeated for each participant. Because several test conditions are tested on each participant, it can lead to fatigue, which can decrease performance, and learning effect, which can increase performance [56]. The learning effect refers to how a participant might learn how to perform a task during the first conditions, which can affect the result of later conditions. To minimize these aspects, counterbalancing should be used when using within-subject design (changing the orders of the test conditions the participants are presented with). A benefit of using a within-subject design is that the variance is similar across all test conditions. The variance refers to differences in people's physical condition, personality, etc. that potentially could affect the results.

Between-subject: If each participant is only tested on one of the test conditions, the study design is between-subject [66]. Because of this, more participants are needed to achieve the same statistical rigour as a within-subject design would require. Individual differences between participants can also affect the performance and results. To mitigate this, the participant groups should be recruited to be as similar as possible. However, a between-subject design limits the learning effect and other interferences between conditions as each participant is only exposed to one test condition [56].

3.4.2 Study Methods

In empirical studies concerning personal informatics, surveys and interviews are two of the most commonly used methods [31]. These methods are either used separately, together, or together with another study method. Other frequently used methods within other fields are focus groups, observations, and case studies, but these are only used sparsely in personal informatics studies.

Survey: Not only popular within personal informatics and HCI, surveys are also one of the most frequently used research methods across all research fields [56]. Used to explain behaviours, describe populations, and gather opinions, surveys are a set of questions participants are asked to answer. Since surveys are generally self-administered, a large number of responses can easily be collected which can make understanding the "big picture" easier. There are a number of available tools that can be used to create surveys, and some popular ones are Google Forms, Qualtrics XM, and SurveyMonkey.

By using a survey, both the time-consuming process and the possible error due to human factors that are present in any manual data collection can be eliminated [82]. In addition, since surveys can be answered at the participant's own convenience, people are more likely to feel relaxed and be in the right state of mind. By making the survey anonymous, participants can be encouraged to respond more honestly and openly.

Vignette study: The vignette study method consists of two elements; a vignette experiment and a traditional survey [4]. By briefly describing a person or situation, vignette experiments can extract participants' judgment on these scenarios. A survey can then be used to collect the participants' responses. Using a vignette experiment together with the methodology of a survey can provide promising results, as it combines the high external validity of a traditional survey with the high internal validity of classic experimental research [1, 4]. Vignettes are also very flexible, as they are not restricted to written format but can also utilize videos, images, and other media.

Vignette studies allow the researcher to exclude factors that can confound the results, and include those that are relevant [1]. Considering the topic of this thesis, in a realistic tracking tool where the studied visualisations are intended to be found, there are naturally a lot more features of interest than what would be feasible for this project. By using the vignette design, the variable of interest can be isolated and variations that can be caused by other features of the tool can be minimized [47].

Interviews: For the purpose of collecting personal accounts of opinions, perceptions, attitudes, and experiences, interviews can be seen as a fundamental research method for firsthand contact with people [15]. As interviews can be structured, unstructured, or semi-structured, the method can be performed in many ways. Structured interviews are easier to analyse, but can be perceived as more stiff and formal. Unstructured interviews on the other hand are more like a conversation, but because of that, they are harder to analyse.

Interviews are a powerful method to involve users in the design process. The method generates deep and qualitative data and can be used to do initial exploration, gather requirements, and obtain evaluation reactions [56]. Interview questions should be open-ended in order to invite the participant to answer in more depth, and they should be unbiased to not guide the answer in a particular direction. To document an interview, notes, audio, or video recording should be used so that the data can be analysed afterwards. One way to analyse interviews is to use Thematic analysis [3]. This method is used to find themes by looking for patterns in the qualitative data, using coding to do so.

3.4.3 Measuring Scales

In HCI experiments, Dependent variables are used to measure human behaviour [66]. These come in many different forms, such as speed, accuracy, key actions, and self-reported answers to scale questions. A few examples of developed scales evaluating different aspects are the Technology-Supported Reflection Inventory (TSRI), the User Engagement Scale (UES), the System Usability Scale (SUS), and the Rumination-Reflection Questionnaire (RRQ). These scales are described in further detail below, and all of them are measured on a 5- or 7-point Likert scale, ranging from Strongly disagree to Strongly agree.

TSRI: The Technology-Supported Reflection Inventory is used to evaluate the effectiveness in which a system supports self-reflection [11]. It allows for the comparison and evaluation of different designs and can be used during the early stages of an interactive system design process. The scale consists of three dimensions; Insight, Exploration, and Comparison. It has 9 items on a 7-point Likert scale with the items being the following;

1. Using the system has led to a wake-up call to make changes in my life.
2. As a result of using the system, I have changed how I approach things.
3. Using the system gives me ideas on how to overcome challenges.
4. I enjoy exploring my data with the system.
5. The system makes it easy to get an overview of my personal data.

6. The system makes it easy to review my long-term personal data.
7. I reflect on my data in the system with others.
8. The system helps me to discuss my data with others.
9. The system makes me think about how my personal data relates with that of others.

RRQ: The Rumination-Reflection Questionnaire can be used to measure rumination and reflection [97]. The RRQ consists of a Reflection Scale, a Rumination Scale, a Trait Reflection Scale, and a Trait Rumination Scale, with all scales using a 5-point Likert scale. The Trait Reflection Scale includes 12 items and evaluates how reflective a person is, which can be based on the extent in which the person has reflective personality traits. The 12 Trait Reflection items are;

1. I love exploring my "inner" self.
2. I often love to look at my life in philosophical ways.
3. I love to meditate on the nature and meaning of things.
4. I don't really care for introspective or self-reflective thinking.
5. My attitudes and feelings about things fascinate me.
6. I love analyzing why I do things.
7. I don't care much for self-analysis.
8. I'm not really a meditative type of person.
9. Philosophical or abstract thinking doesn't appeal to me that much.
10. Contemplating myself isn't my idea of fun.
11. People often say I'm a "deep", introspective type of person.
12. I'm very self-inquisitive by nature.

UES-SF: The User Engagement Scale can be used to measure a user's engagement, a measure of high interest within Human-computer interaction [79]. It has been used in many different digital domains such as online news, education, and social networking systems, to name a few. The original version had 31 items and 6 dimensions but has since been revised to include 30 items and 4 dimensions (Focused Attention, Perceived Usability, Aesthetic Appeal, and Reward Factor). In addition, a short version has been created consisting of 12 items which can be viewed below.

1. I lost myself in this experience.
2. The time I spent using Application X just slipped away.

3. I was absorbed in this experience.
4. I felt frustrated while using this Application X.
5. I found this Application X confusing to use.
6. Using this Application X was taxing.
7. This Application X was attractive.
8. This Application X was aesthetically appealing.
9. This Application X appealed to my senses.
10. Using Application X was worthwhile.
11. My experience was rewarding.
12. I felt interested in this experience.

SUS: The System Usability Scale is a quick and cheap tool that can be used to test usability [14]. The results from the use of this scale are reliable even on small sample sizes, and the scale shows a clear difference between unusable and usable systems, making it a valid tool. The tool has been used to evaluate a wide range of systems and is the most used scale for measuring perceptions of usability. The scale consists of 10 items measured on a 5-point Likert Scale. The 10 items are the following;

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

3.4.4 Study Design Documents

Before designing the visualisations and implementing the application, a Study Design Document was created. The document was created in order to plan the evaluation process as the designs were ultimately created and used in order to explore the research question. By doing this first, no more than the minimal viable product needed to run the study and explore the research question would have to be designed and developed. Many alternatives were considered, but ultimately, I decided to conduct two studies. A short summary of the study designs can be seen below. The full explanation and motivations for why these study designs were chosen are included in the studies' respective chapters.

Study I: Online survey: The chosen study method for the first study was an online survey combined with an experimental vignette. It has a between-subject design with two independent, resulting in six conditions, and two dependent variables, measured using established scales. The six conditions are different levels of abstractness in visualisations combined with goal completion. Hypothetical data will be used to display the visualisations.

Study II: Interviews: Following the first study, a second study will be conducted in the form of interviews. The interviews will be focused on the different versions of the visualisations, presented in an application where participants get to explore them in their intended context. In addition to the interview, the participants will be asked to fill out a short questionnaire measuring the Usability of the application.

4

Design & Implementation

When all the planning, background research, and study design decisions had been made, the practical work could begin. The development of first the visual designs, then the application design, and lastly the implementation of the application are described in this chapter. In addition to this, motivations for all decisions made are presented. The work accomplished and described here refers to the second and third phases of the project, outlined in Chapter 3, *Method*.

4.1 Visual Designs

This section covers the process of creating the designs of the visualisations. First, the base condition was decided before several abstract visualisations were designed. Users were involved during the design process in the form of a two-part pre-study as a way to guide the work. With the participants' opinions in mind, the final abstract designs were established.

4.1.1 Choosing a Base Condition

The visualisation with a low level of abstraction, which could be seen as a base condition, will be in the form of a bar chart. The reasons for choosing the bar chart are because it is a standard representation making most people comfortable in reading them [23], it is a commonly used visualisation in many areas [2] including tracking applications such as Apple Health and the Fitbit app, and many studies within personal informatics have used it as well [34, 59, 74, 96]. Both [34] and [96] used bar charts similar to the one Fitbit has on their website and compared the bar charts against visualisations, with different levels of abstractness, created by the authors.

The design of the bar chart, currently without any colours, can be seen in Figure 4.1. It shows how it looks when the progress toward the goal is around 10%, 50%, and when the goal has been met. The goal is represented similarly to what can be found on the Fitbit dashboard, with an icon metaphor instead of the exact number. A few guiding lines are also included to help the user asses the current progress. The lack of numbers used in this design is to make sure all levels of the condition show the exact same information, albeit in different forms.

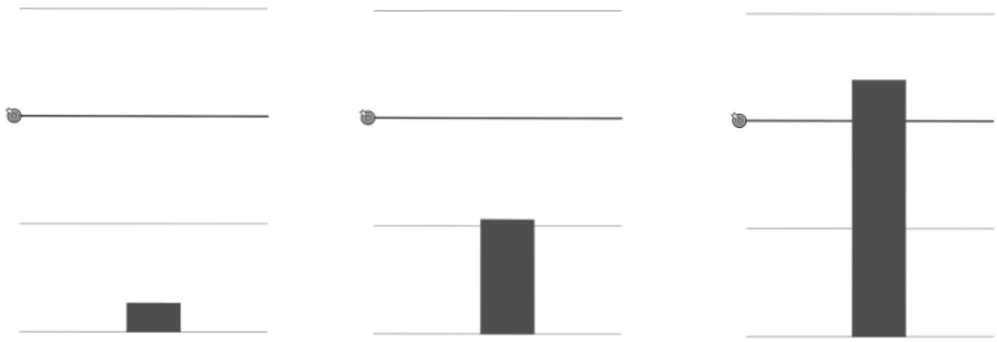


Figure 4.1: The design of the bar chart, used as a base condition.

4.1.2 Creating Abstract Designs

The two levels of medium and high abstraction would need to be designed in an informed but creative way. A number of design principles for behaviour change technologies have in previous work been identified [24], and these will be used as guidelines when creating the designs. The design principles are; Abstract & Reflective, Unobtrusive, Public, Aesthetic, Positive, Controllable, Historical, and Comprehensive. Another study exploring physical abstract representations has devices similar design principles [87]. Their principles are Aesthetics, Abstraction, and Comparison and Correlation, with the Correlation aspect of the last principle being the most unique addition. This principle suggests that if goals are present, they should be integrated into the visual design.

To narrow down the design scope, I decided to focus on goal-based designs, one of the insight types defined following a number of design workshops with professional graphic designers [2]. The motivation for choosing this type of design is because of the importance goals have in personal informatics, with having a goal in mind being the most common reason for starting to track [33]. For a goal to be effective, the progress towards the goal must be presented as feedback to the user [62], which goal-based designs does. It is also often easier to process for the user when feedback is presented as progress in comparison to the target [41], and trackers like to check their data to know their current status towards their goal [58, 86]. Additionally, if a goal has been reached, it should be reflected in the visualisation as well [87].

From studying bullet journal trackers, it was found that individuals often have positive emotional attachments to objects such as geometric shapes, animals, food, or flora [6]. Utilizing this, colouring a drawing of a cat with the mood of the day is an example of a way to track mood in a personalized and creative way. Considering this information, exploring designs that could create a positive emotional attachment will not be neglected.

Brainwriting: A short session of brainwriting was held to generate initial ideas for goal-based designs. Inspiration was taken from existing abstract art and abstract visualisations used in previous personal informatics studies. From this session, a number of ideas were created which could be categorised into two categories; concrete ideas and abstract concepts. The two categories were surprisingly fitting considering the focus of this thesis. Some examples from the concrete ideas are a skyscraper, compass, and hourglass, and from the abstract concepts, a few ones are "Something filling a container", "A mess that takes form", and "Something flora related".

The brainwriting method was used at this stage of the project since the aim was to generate initial ideas quickly and efficiently. The reason this method was used instead of, for example, Crazy 8 was to not limit the number of generated ideas to eight as well as to not spend time sketching ideas that might not meet the requirements anyways, something that might occur when there is time pressure. Another popular idea-generating method is the 6-3-5 method, which wasn't used for to the obvious fact that there were not 6 participants taking part in the ideation process.

Sketching: Based on the generated ideas, sketches of some of the ideas and concepts were drawn which can be seen in Figure 4.2. All ideas show both how they would look before a goal has been met and after a goal has been met, except for b) which has both cases included in its undeveloped design. The colour choices in these sketches were arbitrary and the colours were only used to experiment with how hue, saturation, and brightness, together with form, could show progress.

One issue that blocked many ideas from moving forward and being sketched was how to design them when a goal has been met, one of the important guidelines identified from previous work. Illustrating the user's progress past their goal couldn't be done in an adequate way. More sketches than what can be seen in Figure 4.2 were also not drawn because the ideas started to resemble one another, such as f) and i), and c) and e), so the number of alternatives appeared to be enough at that moment.

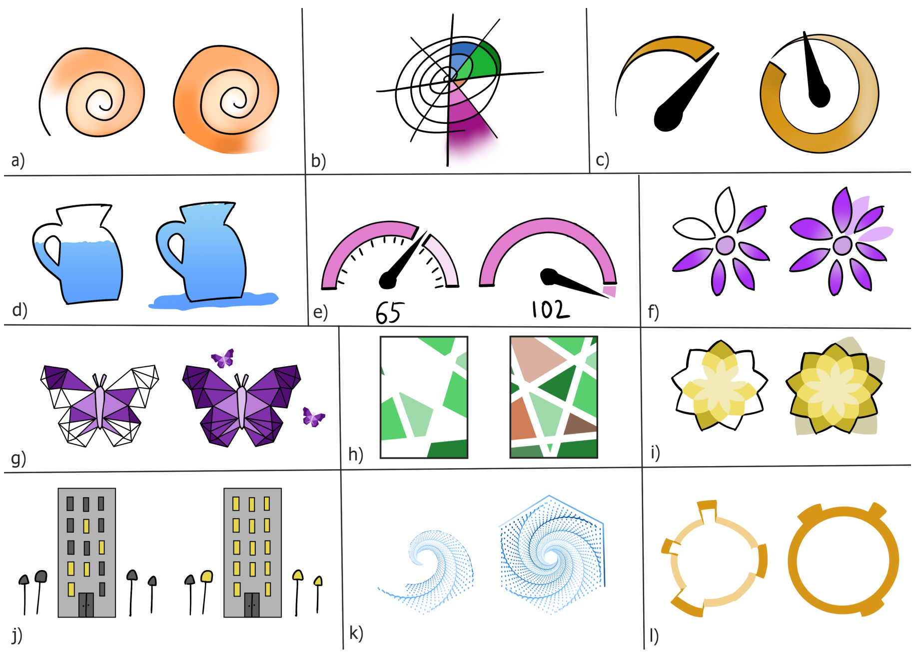


Figure 4.2: The first visualisation designs created out of the ideas generated from brainwriting.

4.1.3 Pre-Study, part I

Involving users in the design process is an important part of creating a useful design that supports users' needs [81]. It is also the main focus of user-centered design, where users are highly involved as a means to create accessible and usable products [37]. Motivated by the importance of user involvement, it was deemed relevant to ask a number of potential users for their opinion on the currently generated design alternatives.

Participants: Eight people took part in the first part of the short pre-study. The participants consisted of 6 females and 2 males, with ages from 20 years old to 58 years old. None of the participants was currently tracking their behaviour, nor had done so previously. Convenience sampling [3] was mainly used to recruit the participants, together with Snowball sampling [3].

Apparatus: Figure 4.2 were shown to the participants who were then asked to select the three most interesting and/or appealing designs, in no particular order. Participants were given no other explanation of what the designs represented or in what context they would be used. This was to avoid bias such as people's inherent opinions on fitness tracking or how visualisations for fitness tracking data should look like. The figure was shown using a mobile phone and the participants' answers, spoken aloud, were written down. The answers were then compiled in order to create a bar chart summarising the results.

Results: Of course, this simple user test could only be used as a guide since the designs had varying levels of refinement. It is also unsure whether people were influenced by the colours, the form, or a mixture of both. Fairly comparing the designs thus becomes troublesome. Nevertheless, Figure 4.3 shows a summary of the participants' choices. As seen from the results, g) was found particularly appealing to the participants, as all participants except one picked that design. Besides that, all other designs were perceived as fairly equal, apart from a) and l) which received no votes. It is possible this could be due to the simplicity of these designs.

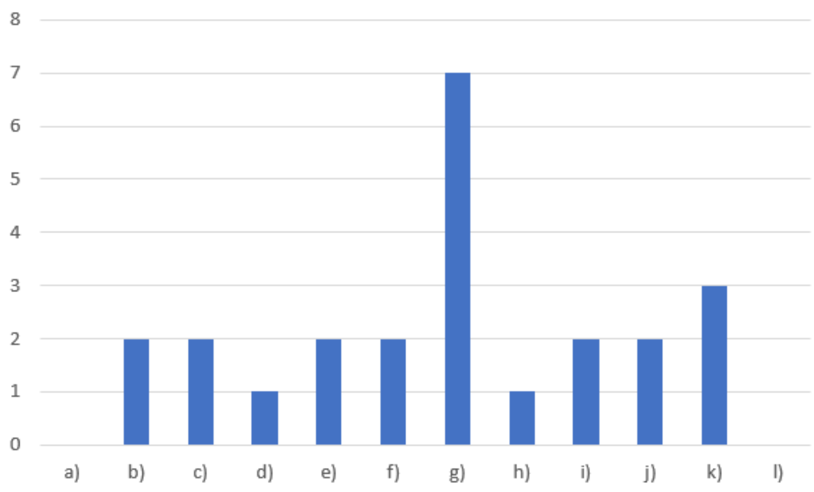


Figure 4.3: Summary of the votes on the initial design sketches.

4.1.4 Refining the Designs

With the results from the short user involvement taken into account together with suggestions from the project's supervisor, I decided to use the butterfly design, g), and to develop it further. Not only was it by far the most popular design in the user test, but it also has the potential to evoke a positive emotional attachment due to its appearance [6]. It was also one of the designs that were based on an existing object rather than being completely abstract, making it easier for the user to relate to.

The next step was two create two versions of the design, one with medium abstraction and one with high abstraction. Abstraction can be defined as the process of eliminating details, removing inessential information, and generalization [99]. With this in mind, the version with medium abstraction was designed to include the segments from the initial sketch. The version with high abstraction was designed to only have an outline to show the goal. By removing the lines creating the segments from the design of medium abstraction, which are not essential but help the user construct a more absolute interpretation, a visualisation of higher abstraction is created.

To show the progress towards the goal, the medium version has the segments filled with solid colours growing randomly while the high version has a circular gradient growing from the middle. Making progress past the goal is visualised in the form of small butterflies next to the large one. Figure 4.4 shows the outlines of the two versions as well as how the visualisations will look if the user has made no progress toward the goal.

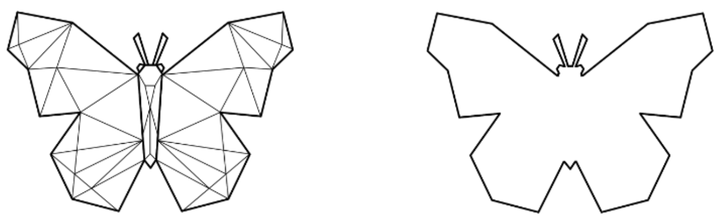


Figure 4.4: The design on the left is the version with Medium abstraction and the design on the right is the version with High abstraction.

Colours: An important aspect of any design is the colour choices. Emotional well-being can be positively affected by viewing colourful representations [6], so including colours was deemed essential. Nevertheless, colours would not be used in a negative manner, such as displaying unmet goals in red and achieved goals in green. This is based on the design principle that technology should be positive [24], and previous research which has shown that negative framing increases ruminations [63, 74].

To avoid the classic positive and negative association that green and red has, those colours were not going to be used altogether. This decision was also made to increase the design's abstractness, as the common association was removed. When trying out different colour combinations, it became clear that by removing green and red, the colour options became limited. Because the colours were shown as a gradient for one of the abstraction levels, the colour palette had to be either monochromatic or analogous as a complementary palette wasn't harmonious. Five alternative colour palettes were created which can be seen in Figure 4.5. Three were monochromatic (a, b, and c), using one hue and only changing brightness and saturation, and two were analogous (d and e), mainly changing in hue.

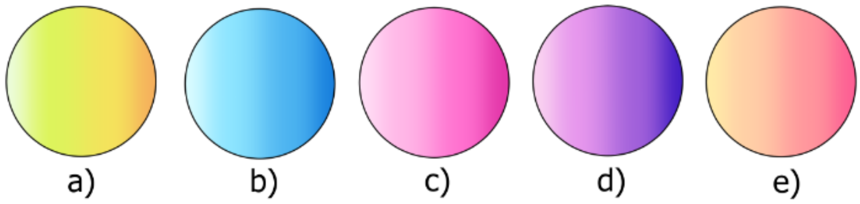


Figure 4.5: The different colour palettes created.

4.1.5 Pre-Study, part II

As part II of the pre-study, users were again involved. This was done in order to get people's opinions on the colour palettes, as colours are equally important as the form of a design. Fewer alternatives were used this time, and because of this, participants were also asked to choose fewer alternatives.

Participants: This time, 10 people (7 females and 3 males) ranging from age 20 to 58 were inquired. Similar to part I of the pre-study, none of the participants were currently or had previously been tracking their behaviour. Participants were recruited using Convenience sampling [3] and some of them were used in part I of the pre-study as well.

Apparatus: Participants were asked to pick the two colour palettes they found most appealing without knowing the context. This was done by showing Figure 4.5 on a mobile phone to the participant or sending the figure to them using a chat application. Answers were spoken or written in response to the inquiry and noted down so that the results could later be compiled.

Results: A summary of the participants' opinions can be seen in Figure 4.6. The palettes were first created as gradients and then explored in the medium abstraction version. When used as single colours in the medium version, it was revealed that monochromatic palettes generally looked more harmonious. However, participants seemed to prefer the analogous palettes in general, with e) being the most voted palette.

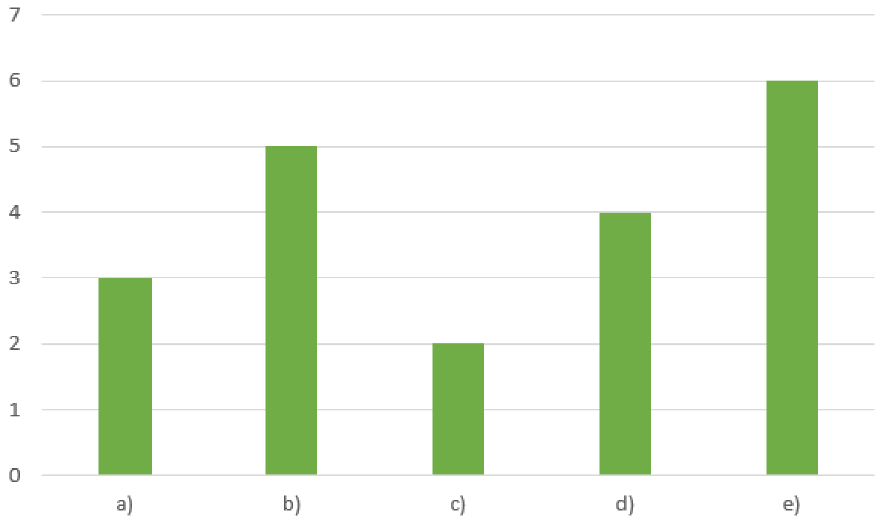


Figure 4.6: Votes on the created colour palettes.

Deciding Colours: The contrast between colours used in an interface is an important aspect of visual accessibility [39]. Elements should have colours with a certain level of contrast to optimize usability for a wide range of users. To test the contrast of a colour palette, a black-and-white filter can be applied. Figure 4.7 shows the created colour palettes as black and white. It was clear that e) lacked significantly in contrast. Since b) was the second most-voted palette and also one of the palettes performing best in terms of contrast, it was chosen to be used for the final design. In addition, because b) has a monochromatic colour palette, it is better suited for people with colour blindness as compared to when several colours are used together. Blue also happens to remain fairly unchanged when viewed by people across a variety of different types of colour blindness.

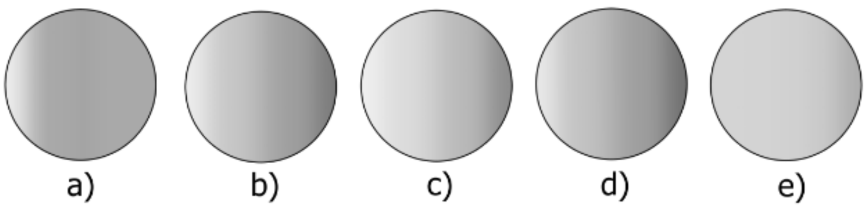


Figure 4.7: Testing the colour palettes as black and white, revealing the contrast between the values.

4.1.6 Final Abstract Designs

The final designs for the medium and high levels of abstraction are shown in Figure 4.8. The figure shows the two designs when the progress toward the goal is around 10%, 50%, and when there has been progress past the goal. The designs fulfil several of the design principles for visualisations and behaviour change technologies presented earlier. The most prominent ones are highlighted below.

- *Aesthetic:* To be meaningful, a visualisation should have a balance of aesthetic and informative properties [87]. Visualisations should also be aesthetically pleasing, regardless of the user's performance. To support a user's personal style, aspects of the technology should be attractive and comfortable [24].
- *Abstract:* The use of abstraction increases privacy [24, 87]. When viewed by others, it is not clear what the visualisation represents. This allows the users to reflect in more situations and view their progress towards the goal [24].
- *Correlation:* Goals are integrated into the visual design [87]. In addition, progress both towards and past the goal is clearly visible. This allows the user to make both relative and absolute estimations.
- *Public:* Data should be presented (and collected) in a way that, if viewed by others, does not make the user uncomfortable [24]. The data should be available at all times and likely will be viewed in common spaces and in the presence of others.
- *Positive:* Positive reinforcement should be used to encourage change [24]. Desired behaviour should be rewarded, but lack of the desired behaviour should not be punished. The use of punishments can result in reduced use of the technology.

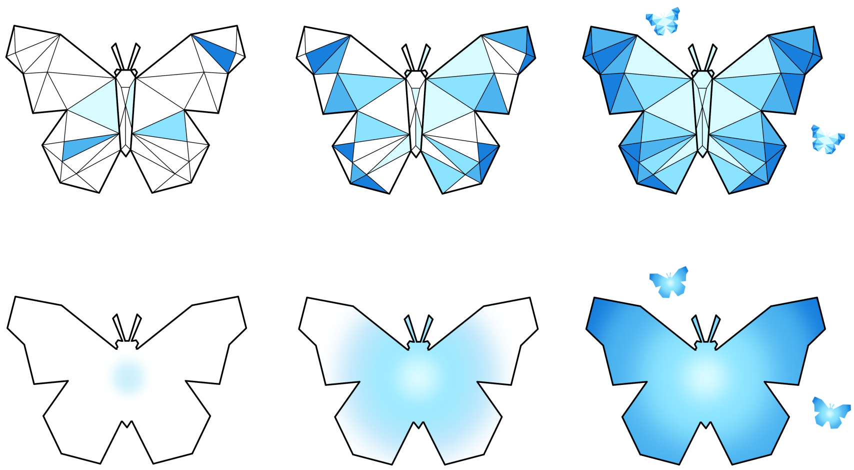


Figure 4.8: The final designs of the visualisations, showing three different progress statuses. The top version has medium abstraction and the bottom version has high abstraction.

4.2 Application Design

When the visualisations were completed, the application was designed. This was first done as a low-fidelity sketch, and then as a high-fidelity prototype. Before that, a few aspects regarding the interface was decided upon, such as navigation and brightness mode. The section ends with design motivations for why the application looks the way it does.

4.2.1 Designing the Interface

Since the focus of the thesis work were visualisations, the simple application created to test the visualisations in the intended context didn't receive as much attention. However, there were still a few design options that needed to be considered. All three visualisations would be shown to the user within the application, allowing the user to switch between them. Three different options for navigating between the visualisations were considered;

- *Single button:* When pressed, the visualisation is switched to the next.
- *Segmented buttons:* Equally many buttons as there are visualisations. Placed together and switches to the chosen visualisation when pressed.
- *Screen Carousel:* The user swipes left and right to access the different visualisations.

A single button takes up less space than segmented buttons but is also less intuitive since the desired visualisation can't always be directly switched to (since there are three of them). Both segmented buttons and the screen carousel offer the possibility to navigate to the desired view directly if the screen carousel wraps around from the end to the beginning. The screen carousel requires the user to remember where the items are placed though, if no mistakes are to be made.

Most mobile applications have a transient posture, which means it is only used for short moments with a focus on particular tasks [25]. For that reason, an application should be clear and simple. Motivated by this, I decided to use the screen carousel since that would make the design more clear and simple with no buttons having to use screen estate.

The next consideration was deciding between light mode and dark mode for the application. Naturally, both options should be available in order to accommodate all people, but for the small user test in this thesis, it was necessary to have the application be consistent. The choice landed on dark mode. Dark mode colour schemes can reduce visual fatigue and increase visual acuity [51]. It is also often preferred by users and decreases battery consumption in devices with OLED displays [27], which is the case for most mobile devices released in recent years.

4.2.2 Low-Fidelity Sketch

The first created paper sketch of the application can be seen in Figure 4.9. This was done before the visualisation designs were explored and developed, but the general idea was still relevant at this point in time. The first view has a text label stating "Set today's step goal" and a text field where the user can enter his goal for the day. If unsure, the user can press the suggestion button and a suggested goal will be provided. When confirming what goal to use for the day, the user is navigated to the second view. This view has different visualisations, where the user can change between them by swiping.

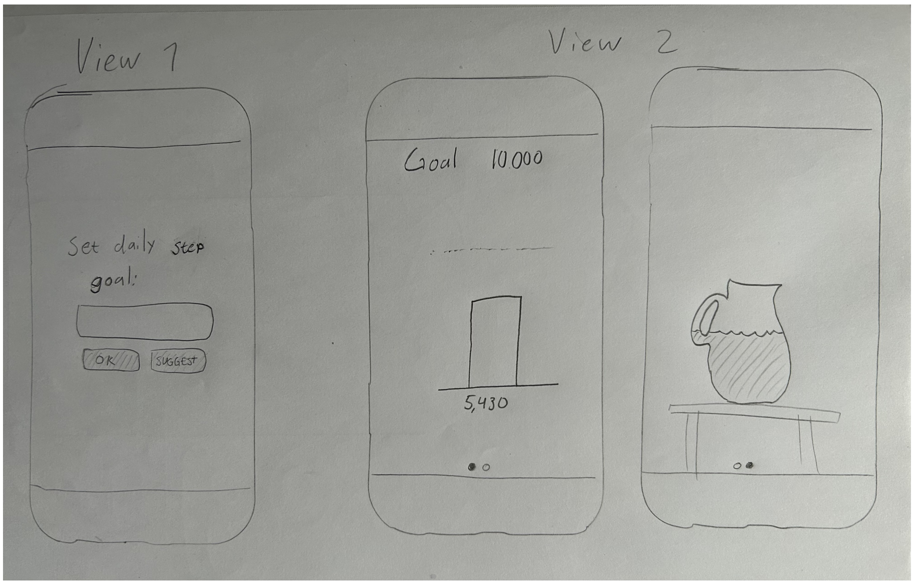


Figure 4.9: The first sketches of the different views of the applications. View 1 lets the user set a daily goal while view 2 shows the different visualisations.

4.2.3 High-Fidelity Prototype

Based on the simple sketch, a high-fidelity prototype was created to use as a guide for the implementation. This was done using Figma, a collaborative interface design tool which can be found here; *https://www.figma.com/*. Figure 4.10 shows the four views. The colour palette created for the visualisations was used for the whole application as well, to achieve a consistent and uniform look.

Both the buttons and the text field have rounded corners. There are several reasons for using round corners in general rather than sharp edges. The smoother feel round corners provide can evoke an emotional connection towards the device and viewing rounded rectangles requires less cognitive load [89]. The round corners also make the design appear more friendly, which can increase the feeling of warmth and trust.

The two buttons have different designs with the primary action, the confirm button, having high emphasise while the optional action, the suggestion button, having low emphasise. This is in accordance with Material Design's guidelines [39] for how buttons should be designed, where each screen should only have one prominent button and the other buttons should clearly communicate their lesser importance.

There are two types of text fields, filled and outlined. For this design, the outlined one was chosen. Both types of text fields provide the same functionality and deciding between the two should thus be based on what works best with current the visual design as well as being distinct from surrounding content [39]. The text field has a label placed in the middle of the field, as all text fields should have a label either inside or close to the field.

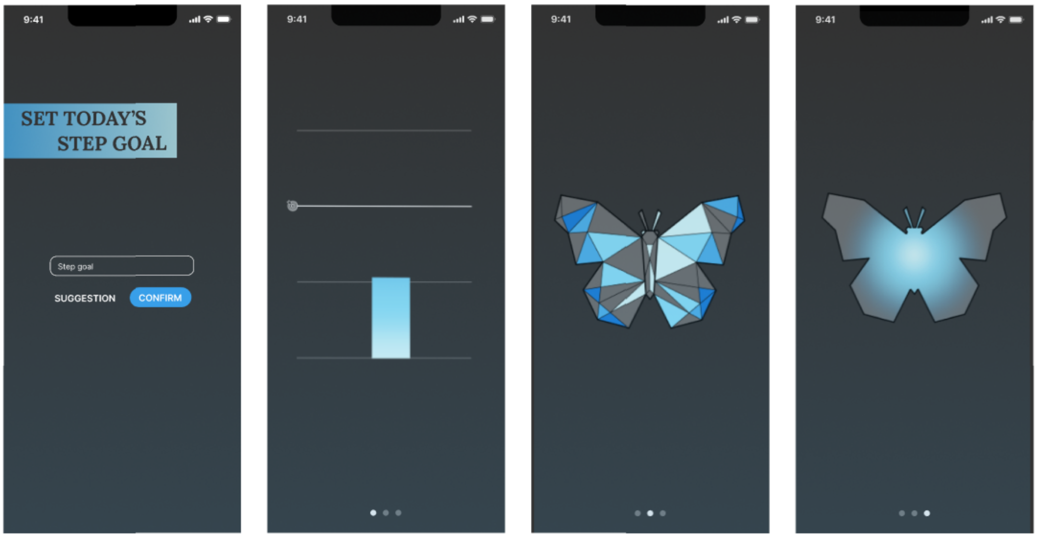


Figure 4.10: The high-fidelity prototype created in Figma. First a screen for setting the goal and then the screens with the visualisations.

4.2.4 Design Motivations

Several motivations for the different design choices have already been presented previously. This section covers the remaining areas with a focus on data from a single day, the use of several visualisations together, goals, and the interface itself.

Single-day Data: One of the guidelines for behaviour change technologies is Historical, meaning that information regarding a user's past behaviour should be accessible and reasonable [24]. While some trackers, especially long-term, like to look for patterns over time [58] where the important things are usually found [20], it turns out that some people are not interested in historic data [40, 86]. Visualisations showing the day-to-day progress are often preferred by short-term trackers [32], and people, in general, are often influenced by day-to-day activities and single data points [20].

Because both studies will run during one day and not as a field study, data from a single day will be used. Since several previous studies have acknowledged the effect day-to-day progress has on people, studying visualisations with one day's worth of data was believed to be interesting and promising.

Several Visualisations: While only one of the visualisations was going to be presented to each participant during the first study, the second study would allow the participants to explore all of them. The reason for this is because of the many benefits the use of several visualisations can provide.

First, technologies aimed at changing behaviours should have the option to change between visualisations so that people can change the interface based on the context and setting, increasing the opportunities for reflection [24]. Secondly, users *want* to be able to choose between visualisations [80]. People have different preferences in terms of how data should be visualised, and by combining several, the diversity in people's needs can be addressed [96]. Including more than one visualisation may also allow data to be presented in a more meaningful way [96] and help the user see the full picture [30]. Lastly, by providing both abstract art and charts, both engagement and enjoyment can be increased over time [34, 96].

Goals: Overall, the application is designed to be more reflective than persuasive [70]. Therefore, the data is displayed in a neutral way and the application is intended to support the users towards their self-chosen goals. Users also tend to prefer self-set goals over assigned or pre-decided goals [22]. This is partly because people need different goals to be appropriate. One goal is not ideal for all users and a system should take the user's physical abilities into account when it comes to goals [34]. A factor that often hampers long-term engagement is when the system goals do not match people's goals [43].

However, novice users have limited knowledge regarding their daily physical activity [41] and tend to not know how much they walk, for example [40]. This is why it is essential to include the possibility to have a goal suggested when self-set goals are used. The suggested goal should be an appropriate number and not too difficult to achieve. If the goal can be achieved, it can motivate users to reflect and challenge themselves [40]. The combination of self-set goals and suggestions has been used in a previous study [40].

Having the user set a daily goal has been used as well [41]. By using daily goals, the application becomes more flexible. Some days it might not be possible to reach one's optimal goal. Then it might be more comfortable and motivating for the user to set an appropriate goal that can be achieved rather than knowing that goal failure is inevitable. Illness and work deadlines are some factors that might hinder a person from temporarily achieving her daily goal [22]. Goals also change over time. An initial goal might have been too ambitious, or it might become too "easy" once a person's fitness level has improved [43]. This change is something many tools don't account for.

Interface: Although the design of the application is very simple, the interaction design principle presented in Chapter 2, *Related Work*, were taken into account when designing the few aspects present. From the principles presented by Sharp et al., the most applicable are;

- *Feedback:* When the user is navigated to the second view, and when the suggestion dialogue box appears, this is done without delay.
- *Consistency:* The design is consistent across all views, using the same style and colour palette.
- *Affordance:* The primary button has the look of a button, according to today's standards. Previously, metaphors were widely used in interfaces [25], such as

giving buttons drop shadows. Nowadays, interfaces rely more on idioms by using simple visual and behaviour idioms which are easy to learn. This creates cleaner interfaces which cause less visual excise.

Included in the Conceptual Principles are that computers should be considerate and "act like humans" [25]. A considerate product is self-confident. Dialogue boxes asking "Are you sure?" when an action is performed should not exist, and no such dialogue will appear from pressing the confirm button in this application.

Technologies should have harmonious interactions [25]. There are no universal rules as to what creates harmonious interaction, but one aspect that doesn't bother the user's flow is the notion "less is more". With the use of a minimalist approach to a product's design, the purpose of the tool is clearly understood. The number of elements in an interface should be reduced without affecting the product's capability or making it more difficult to use. Another aspect is to provide choices rather than questions [25]. Rather than having the question "Do you wish to set your daily goal yourself or have a goal suggested to you?" on the first view, the user has the optional suggestion button.

Information should be contextualized. This means that instead of showing the exact numbers, visualisations should be used. The use of visualisations makes it easier to comprehend the proportion and scale of the numbers being displayed. This principle is also another motivation for using a bar chart instead of, for example, a text saying "Step count: 6 789 / 10 000" which can be found in some current tracking tools. The information presented in both options is the same, but the exact number is often not really relevant to the user and should therefore not be the main focus.

As for Behaviour Principles, one aspect of a good visual interface design is to keep it simple [25]. This refers to having a restricted colour palette, preferably using sans-serif fonts, especially for small texts, and only using one or two typefaces in the interface. All of these aspects are fulfilled in the design of the application.

4.3 Implementation

In this section, the process of implementing the designed application is covered. A technical review is done in order to make sure the plans are feasible before doing further work. After that, details about the code, creating several versions of the visual design, and testing the code are covered.

4.3.1 Technical Review

I deemed it important to research the technical possibilities before going further into the practical work of this thesis. First, I decided to use Flutter for implementing the application. Flutter is created by Google and is a free open-source framework. With a single codebase, applications can be created for mobile devices with both iOS and Android as the operating system as well as desktop, web, and embedded devices. Because of the framework's flexibility, and because it can be used together with Material Design which has accessibility standards integrated [39], Flutter was chosen as the framework over other options for mobile development such as React Native, jQuery Mobile, Native Scripts, and Ionic.

The documentation for both Flutter and Material Design was reviewed in order to facilitate the upcoming implementation process. For Flutter, widgets were the main focus which are the building blocks used to build the application's user interface. The Flutter documentation can be found here: *https://docs.flutter.dev/*. For Material Design, the focus was to check the different design opportunities as well as get some design inspiration. The Material Design documentation can be found here: *https://m3.material.io/*.

Based on the decision to use Flutter for the development, two main questions had to be addressed before moving forward. This was to make sure it was feasible to do the intended application development and also conduct the planned studies. The two questions were the following;

- How to collect tracked data from a phone?
- How to use the exported data?

Collect data: Despite Flutter's flexibility, it was necessary to focus on one device so that the study could be consistent amongst the participants. Since Apple currently is the leading mobile phone brand [61], and because of my experience and knowledge of iPhones, I decided to focus on iPhones. iPhones have a Health app preinstalled on all devices running iOS 8 or later. This app contains several types of health data and also tracks physical activity data such as walking distance and step count. Exporting health data can easily be done within the app.

Use data: Using the data proved to be more difficult. The Health app data is exported in XML format (eXtensible Markup Language), a tool used to store and transport data independent of hardware and software. This format is not easy to read or use, with the export test generating a file with over 300 000 lines of information. However, using an online converter, the file could be converted into CSV format (Comma-separated values). This format could be imported into a Flutter project and then parsed so that the information could be used. The following website was used to convert the XML file to CSV files; *https://www.ericwolter.com/projects/apple-health-export/*.

4.3.2 Creating Visuals

In a real application, visualisations are dynamic and constantly change based on the user's activity. Considering the scope of this thesis, making the visualisations dynamic for the application was neither feasible nor necessary. Since the study using the visualisation would both be sedentary and take no longer than 30 minutes, the participants would not increase their step count during that time. For that reason, it was only necessary that the visualisation appeared to accurately represent the participant's current progress.

A total of 13 versions of each visualisation were created showing different levels of progress. The mapping was based on the relative progress towards and past the goal, with each visualisation version representing a 10% interval. The intervals were the following; 0 - 9%, 10 - 19 %, [...], 100 - 109%, 110 - 119 %, and 120% -. The last interval covers any progress from 120% of the goal completed to infinity.

The visualisations were created using Clip Studio Paint, and the same program was used during the design phase as well. All of the visualisation versions can be seen in Appendix A. The version with low abstraction had the lines changed to black for the appendix in order to be visible.

4.3.3 Writing the Code

With all the visualisation versions created, the code could be written. The code was continuously run through an emulator so it was possible to check that the application looked like it was supposed to. The tools and software used during the implementation were the following;

- *Code editor*: Visual Studio Code
- *Programming language*: Dart
- *Framework*: Flutter
- *Emulator software*: Android Studio
- *Emulator model*: Google Nexus 6

To achieve the different emphasises on the buttons motivated in the section 4.2, *Application Design*, the components called "ElevatedButton" and "TextButton" from Material Design were used. These, together with the TextField component, were adjusted in terms of shapes and colours in order to resemble the look decided upon in the high-fidelity prototype.

Pressing the suggestion button displays a dialogue with text suggesting that 7000 steps is an appropriate step goal to start with. The reason for using 7000 steps as the suggestion was to not have a goal with too high difficulty [40], but at the same time could be perceived as difficult enough so that a greater effort could be put into it [62]. In addition, the classic goal of 10 000 steps per day is generally not necessary. A study showed that the benefits of additional walking were slight after reaching 8000 steps [84]. It is also often unlikely that people reach 10 000 steps per day, especially amongst new trackers, considering how most people in Western nations on average take 5000 steps per day. Adding a few thousand steps per day is not only achievable but also greatly increases a person's health. Figure 4.11 shows the suggestion box.

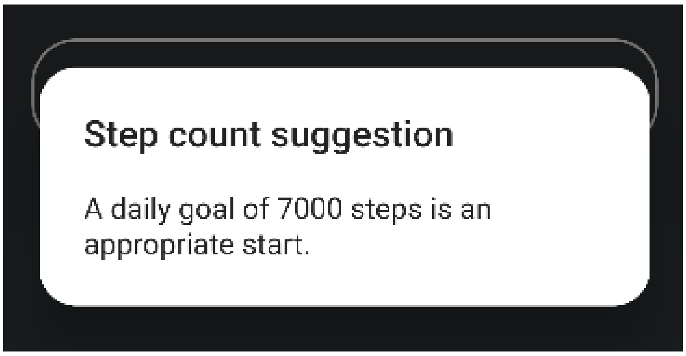


Figure 4.11: The dialogue showing the step goal suggestion.

Confirming the set goal takes the user to the next view, the visualisations. The application design does not allow this action to be reverted, meaning that the user can not change the daily step goal once it has been set. Less reflection might be the result if users are allowed to change the goal once they have seen that they are currently far from achieving it, for example. Since the goal is also daily anyway, the user can change the goal the following day if the goal he set was too difficult.

Because of the emulator used, it is still possible to go back and change the goal when exploring the application. No action was made to prevent this from occurring since this aspect would be beneficial for the studies. Allowing the participants to test different goals would open the possibility for them to give more complete feedback on the visualisations and the changes occurring following the user's progress.

Packages: The visualisation view required a number of Dart packages, to decrease the code needed to be written. Packages can be found on *https://pub.dev/*, and the following packages were used;

- *CSV*: Converts CSV files to lists of values.
- *Page view dot indicator*: Dots indicating which page the user is on.
- *Flutter Spinkit*: Animated loading indicators.

A loading screen is shown to the user before the visualisations can be seen, with the "Wave" animation from the Flutter Spinkit package shown. This is because extracting the information from the CSV file using the CSV package takes a few seconds. When the information has been extracted, the view with the screen carousel is shown. The PageView widget from Flutter was used to create the screen carousel. Each view consists of one of the visualisation versions. This is also where the Page view dot indicator package was used, to indicate which page the user is currently on.

4.3.4 Code Testing

During the implementation process, the application was checked for errors. This was not done more extensively than making sure everything was working as intended and that even if the user does something unintentionally, the program should not break.

To avoid input errors, a few measures were taken. The input length was restricted to 5 characters since allowing for goals over 99 999 steps was deemed unnecessary. The value is checked to make sure it is indeed a number and not a word, and also if it is above 0. If any of these aspects are not true, or if no value has been written, the button will not navigate the user forward.

Not many major errors occurred during the coding process. Most were small errors such as the components not appearing where they should at first. The most notable error was that when parsing the CSV, the wrong information was fetched. This was due to the different ways CSV can be built up. For example, both "," and ";" can be used to separate the items, and different characters can be used to mark the end-of-lines. There is no Internet standard for how CSVs should be formatted. If the values used in a particular file are not specified, the function simply uses the default values. This error was solved quite easily once the source of the problem had been identified.

4.3.5 Final Application

Although very simplistic, the final application is intended for users new to fitness tracking. It allows the user to set a daily goal and view her progress through different visualisations where the goal is visibly integrated. Because of the lack of features, it shouldn't be seen as a final product but rather as a building block for possible further development. With that said, I have decided to call the application FlyFit and the different views can be seen in Figure 4.12.

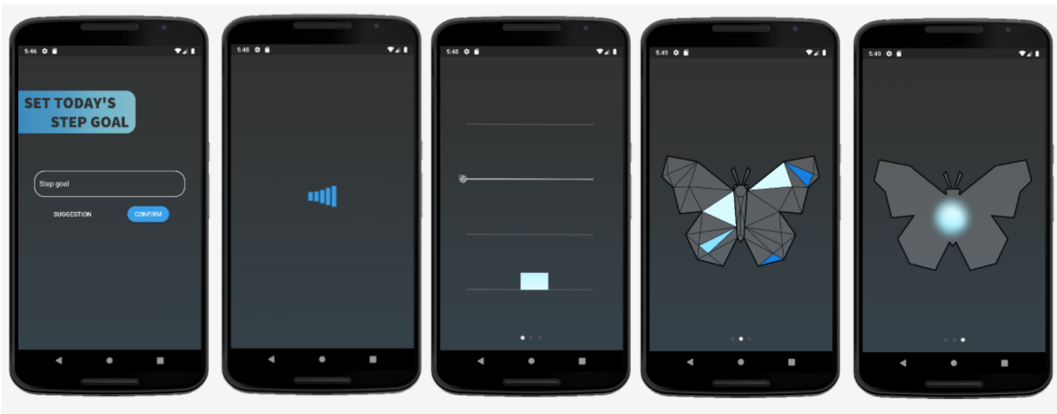


Figure 4.12: All screens of the complete application. The visualisations represent 392 steps taken with 3000 steps as the goal.

5

Study I: Online Survey

The first study, which is part I of the two-part evaluation process, was an online survey using an experimental vignette. The purpose of this study was to explore the levels of abstraction with a large number of participant, in order to get quantitative feedback in terms of self-reflection and engagement.

5.1 Study Design

I decided to use an experimental vignette study design for the study. Besides the combination of high internal and external validity this study design offers, it was also chosen because of its usefulness to gather evidence of causation by having control over the independent variable [1]. Vignette studies have efficiently been used in previous personal informatics studies [12, 74].

Vignette studies use surveys, which by themselves are highly useful. As mentioned, surveys are often used within personal informatics studies, and a few examples of studies using online surveys, in particular, are [10, 33, 46, 57]. By using online surveys, an increasingly larger number of participants could be used compared to what would have been possible if a researcher had to be present at every user test. With more participants, the probability of achieving statistically significant results also increases [66].

Next, I chose to use a between-subject design. Although within-subject design is the most frequently used design within HCI [47, 56], between-subject design is more often used within personal informatics, which can be seen from the many studies that have used it [12, 21, 52, 59, 74, 96]. The reason for this is because of the content personal informatics deals with. Once a person has seen their data, it can't be unseen. The initial experience and reflection occurring from seeing one's data will be vastly different for a second condition. Problems that cannot use within-subject design to accurately investigate them should use between-subject design [56]. Besides the mentioned points, because a between-subject design was chosen, there was no need to take learning effect or fatigue into account.

The first independent variable I decided to use for this study was *Abstraction* with three levels; *Low*, *Medium*, and *High*. Initially, the plan was to only use two levels; concrete and abstract, but doing so can be questionable. By using more than two conditions, the validity of the result can be increased in terms of the key ideas of the research questions [47]. Additionally, even if no difference is found between the levels of the independent variable, interesting data regarding the differences between the construct instances can still be generated. The second independent variable used was *Goal Completion* with two levels; *Completed* and *Not Completed*. People tend to reflect and engage to different extents depending on if they have completed or not completed their goal [40]. It was therefore of interest to include this condition as well.

In the following, the conditions are called; *LowC*, *LowNC*, *MediumC*, *MediumNC*, *HighC*, and *HighNC*, where C refers to Completed and NC refers to Not Completed. The assignment of the six conditions, and which condition the participant was presented with, was randomized. By using randomization, the experiment's internal validity is increased since unknown factors of no interest that influence the experiment are evened out between conditions [47]. The consent form and the vignette used in the survey can be seen in Appendix B, as well as the ending of the survey with the open question.

5.2 Pilot study

A pilot study was run using 12 participants, with two participants per condition. This was done to reveal any errors or unclear points with the survey, before running it with a large number of participants. The open question at the end of the study revealed an issue with the study design. The two participants with condition *HighC* left the following comments:

> *There was only images of blue butterfly instead of visualisations.* (P4, *HighC*)
>
> *I didn't understand the meaning of the butterflies.* (P8, *HighC*)

It appeared that the lack of context made it difficult to understand high abstraction when the goal was completed, as the only thing the participants saw was a larger and a smaller butterfly. The participant leaving the first comment also considered himself experienced with tracking, and the comment could thus be the result of bias from his previous experience with tracking visualisations, where visualisations mostly are concrete.

Because of the comments, adjustments were made to the images shown to the participants. Each condition was changed to include three different stages of the visualisation together with a timestamp so that the participant could see how the visualisation progressed throughout the hypothetical day. Figure 5.1 shows one of the conditions with the updated look, and all conditions can be seen in Appendix C.

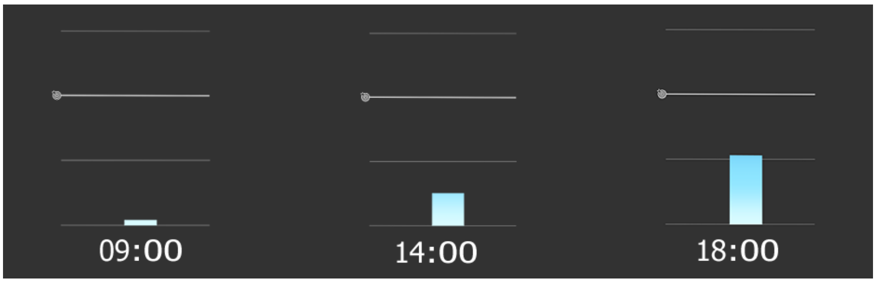


Figure 5.1: Condition *LowNC*, Abstraction: *Low* and Goal Completion: *Not Completed*.

5.3 Participants

Participants were recruited through Prolific, *https://www.prolific.co/*, a website used to conduct online research with high quality. By using rigorous checks, unbiased screening, and fair payments, Prolific assures participants are who they say they are and responses becomes thoughtful and honest. Amazon Mechanical Turk, *https://www.mturk.com/*, is another commonly used website for the same purpose, but it requires more additional filters in order to receive responses of similar quality to Prolific.

A few criteria for the participants were defined beforehand. All participants had to be at least 18 years old, as research on children and young people is problematic from an ethical perspective. Since there is a limit to their ability to assess consequences, obtaining acceptable informed consent is difficult. Next, as the study was conducted in English, participants had to know English well enough to understand the content of the study. Participants also needed to be, for the most part, relatively new to tracking and analysing data. The reason for this was, as mentioned previously, because short-term trackers tend to prefer visualisations showing their day-to-day use [32], and data from a single day were going to be used in the study. Lastly, participants had to not be experts in analysing visualisations considering how the average tracker is not. One of the aims was after all that the visualisations should work well in the personal context, a context where people are generally unfamiliar with how to analyse visualisations.

In order to generalise results, the participants recruited should have the same variability as what is present in the user group that is being studied [47, 66, 67]. For this reason, I chose to use a balanced sample so that the distribution between males and females would be equal. The number of participants tested must also be sufficient [66] and should have characteristics relevant to investigate the research question [47, 68].

A total of n = 222 participants were recruited for the study, with ages from 18 to 79, M = 30.29, SD = 10.32. Out of the participants, 111 were male and 111 were female. The participants received £ 0.75 as compensation for participating in the survey which had an average completion time of 3 minutes 43 seconds. The participants resided in Western countries, mainly the United States, Canada, or Europe. The participants responded if they considered themselves experienced with tracking on a 5-point Likert scale from Strongly disagree to Strongly agree. The mean value for the tracking experience was 3.45 (SD = 1.17), where 3 refers to the middle value with the textual answer being "Neither agree nor disagree".

5.4 Apparatus

Since the study was using an online survey format, the only material the participants interacted with was the survey itself. The survey contained all necessary information, descriptions, and visualisations. All participants answered the same questionnaire questions, with the only difference being which condition they were randomly assigned to.

Data was collected through the survey created with the Qualtrics XM platform, found here; *https://www.qualtrics.com/*. This tool was chosen because of the many features it offers. A large number of available templates, several question types, and facilitating randomly assigning conditions to participants, to mention a few. The survey contained a short consent form, a section about possible influential factors, and a section measuring the participant's trait reflection. The participants were then assigned to a condition before reaching the following two sections of the survey, measuring the dependent variables reflection and engagement respectively. All participants regardless of condition were presented with an optional open question at the end of the survey.

5.4.1 Reflection Scale

Most studies within personal informatics focus on how to create tools that effectively support self-reflection and thus support changing behaviours in a long-lasting way [31]. Because of the importance it has for trackers' success, it was necessary to measure reflection. The scale used to measure reflection was the Technology-Supported Reflection Inventory (TSRI) [11]. The TSRI is artefact-centric and is intended to facilitate researchers to compare designs against each other in terms of their ability to elicit reflection in users. When developing the scale, it was tested using Fitbit dashboards, i.e. fitness data visualisations, making it appropriate for this study. It has also been used in a number of works on abstractness and tracking data [12, 46, 63].

In addition to measuring reflection, trait reflection was measured using parts of the Rumination-Reflection Questionnaire (RRQ) [97], as suggested when using the TSRI [11]. This is to analyse the participants' individual differences, as reflection is both due to people's reflective personality traits and the current state. The difference in which different visualisations can elicit reflection in users cannot be adequately assessed unless trait reflection has been controlled and used as a covariate for the analysis. Previous works have used this scale as well [12, 63, 74].

Other options for reflection scales are the Groningen Reflection Ability Scale (GRAS)

[5] and the Self-Reflection and Insight Scale (SRIS) [42]. Both the SRIS and GRAS measure the reflective capacities in a person and are based on the Psychology insight that reflection is tied to a person's personality [11]. However, interactive technologies can influence the user's reflection capacity as well, and certain qualities affect an interactive technology's ability to support reflection. These reasons together with the fact that the TSRI is artefact-centric are the motivation for choosing said scale to measure reflection.

5.4.2 Engagement Scale

Another important aspect of personal informatics is engagement. As many current systems struggle with engaging users [43], it was relevant to measure the effect the different visualisations have on user engagement. To measure engagement, the short form of the User Engagement Scale (UES) was used [79]. Studies evaluating digital technologies with Western adult populations have been used during the development and testing of the UES. Considering how both the same user group and domain are used in this study, the scale was deemed fitting.

Another widely used engagement scale is the Utrecht Work Engagement Scale (UWES) [88]. However, this scale focuses on measuring engagement within the work environment, aimed at students and employees. The UES on the other hand has been used in many different digital domains [79] as well as in work focused on ambiguity [46].

5.5 Procedure

The participants start the survey by themselves through Prolific and are first presented with a text describing the purpose of the study. The text also describes what will happen in the study, what is expected of the participants, and that the answers will be anonymised. Participants are informed that the study will take approximately 5 minutes to complete and that it is voluntary with participants having the right to withdraw at any moment. This information together with acknowledging being over 18 years old must be consented to for the survey to continue. If it is not consented to, the survey ends.

The next section of the survey included a field where participants entered their Prolific ID so that their demographic data could be recorded, and a question regarding if they consider themselves experienced with fitness tracking. After that, participants' trait reflection is measured using the 12 items from the RRQ [97].

Participants are then randomly assigned to one of the six conditions; *LowC*, *LowNC*, *MediumC*, *MediumNC*, *HighC*, or *HighNC*. Regardless of the condition, the participants were presented with a short vignette, describing a scenario in which the visualisation could be found and used. With the perspective of the person in the scenario and the visualisation in mind, participants were asked to answer the 9 items from the TSRI [11] to measure reflection and then the 12 items from the UES-SF [79] to measure engagement.

The survey ends with an open and optional question asking participants if they have any additional comments or questions regarding the visualisation they were presented with. The last page thanks the participants for their time and redirect them back to Prolific.

6

Study II: Interviews

The second study was conducted as semi-structured interviews. The purpose was to explore the visualisations in the intended context in order to gather deep qualitative insights. The interview questions were based on the results from Study I. The visualisations were explored using FlyFit and the application was in turn assessed in terms of usability.

6.1 Study Design

The interviews were semi-structured to allow for additional questions when necessary while still keeping an overall structure to not make the process of analysing too complicated. Semi-structured interviews tend to be more appropriate when the goal is to dig deeper and search for critical comments, while structured interviews work best when there is a need to compare responses [56].

The reason behind my decision to conduct interviews following the survey was that interviews are useful for collecting detailed qualitative data. While surveys are good for gathering many people's opinions through the large number of participants that can easily be used, the data tend to be limited and "shallow" [56]. Interviews were also chosen as the follow-up method rather than, for example, a focus group or a case study, because of the method's wide use within the field [31], present in many studies [2, 21, 34, 58].

The semi-structured interviews were based around FlyFit with the visualisations included. It was also combined with a short questionnaire, used to measure the usability of the application. By combining the interviews with an artefact, it became more productive [15]. It is also a common strategy to combine interviews with another method such as a questionnaire since collected data can be verified through the use of other means. The interview protocol can be seen in Appendix D.

6.2 Participants

Convenience sampling [3] was used to recruit most of the participants, meaning that available people nearby were recruited. Snowball sampling [3] was used to recruit one of the participants, meaning that a participant recruited another participant in turn. These sampling methods are fast, efficient, and easy to use, but often result in some bias making the results non-generalizable for most aspects. However, since the results from Study I would be the main basis used to answer the research question and the purpose of this study was mainly to enrich those results, I believed these sampling methods were sufficient.

The requirements for the participants were the same as in the previous study. The only addition was that participants also needed to own an iPhone since the application was made to handle data extracted from the Apple Health application. A total of n = 9 participants were recruited for the study, with ages from 21 to 58, M = 37.11, SD = 15.87. Out of the participants, 2 were male and 7 were female.

I believed that using nine participants for the study would be sufficient. Since there are no general guidelines for how many participants should be used in an interview study, the best direction is to look at other studies within the field [56]. Between 10-12 participants have been used in a few other studies with interviews [2, 12, 41, 57]. For the usability part of the study, nine participants were enough as well. Not only because the scale used has shown to be effective on very small samples [14], but also because usability evaluations in general only require a small number of participants in order to reveal most problems in an interface [66].

6.3 Apparatus

The interviews were conducted as one-on-one sessions lasting for approximately 10 minutes, with the shortest taking 5 minutes 49 seconds and the longest lasting for 17 minutes 49 seconds. The audio was recorded using a phone and the recordings were only used to transcribe the interviews afterwards. A HP omen laptop was used to run the application, allowing the participants to try FlyFit on an emulator. The same laptop was also used to let the participants fill out the short questionnaire.

Data was collected through audio recording and the questionnaire. The questionnaire was created using Google Forms with only a section for measuring usability. Demographic data was collected orally. Due to the simplicity of this questionnaire, and because the interviews were the main focus, Google Forms was considered a sufficient tool.

6.3.1 Usability Scale

Testing the usability was simply to assure that, even if the application was simple, poor usability didn't affect the results or took away the focus from the relevant topic. Usability is also an important aspect of interaction design, making it an appropriate measure. To measure usability, the System Usability Scale (SUS) was used. It is a widely used scale [8] with both high reliability and validity [14], which are the motivations for why it was chosen for this study. Since the purpose was only to check the usability quickly, it was considered unnecessary to have an elaborate scale such as the 50 items Software Usability Measurement Inventory scale (SUMI) [53].

6.4 Procedure

Participants were asked if they consented to take part in the interview and to have the audio recording, being assured that the audio recording would only be used to transcribe the responses and that their participation would be anonymous. The participants were also introduced to the topic and approximately how long the interview would take. If these aspects were consented to, the participants were asked to send their current step count data. This was imported into FlyFit and the participants were then allowed to try the application freely. The participants first inputted a step goal into the application and then examined how their own step count status was visualised in terms of their progress towards their set goal. Participants were allowed to go back and change the goal for the interviews, so they could see how the visualisation changed depending on how close or far their current step count was from the goal they set.

When the participants appeared satisfied with trying the application, they were asked to fill in the questionnaire regarding usability. After the questionnaire had been completed, the audio recording started and the first interview question from the scripts was asked. Participants were encouraged to continue using the application during the interview and to answer the questions in as detailed as possible. When it was deemed necessary, follow-up questions were asked as the interviews were semi-structured.

The interview questions were divided into three categories; Comparison, Design, and Interpretation. A total of 15 questions were prepared for the interview script, including some optional which were dependent on the participants' progress shown in the visualisations and how they responded to the other questions. The questions were about how participants experienced the visualisations and the designs, focusing on both shape and colour. There were also some questions to judge how well participants could interpret their results from the visualisations. The interview ended with participants being thanked for participating and asked if they had any additional comments or questions.

7

Results

This section covers the results of the two studies. The results from study I will be presented first, including an analysis of the reflection and engagement scores, and a summary of the comments received through the optional open question. After that, the results from Study II will be presented, which include the usability scores and an analysis of the responses during the interviews.

7.1 Study I Results

The results from Study II are divided into two quantitative parts and one qualitative part. The quantitative parts consist of the results from measuring reflection and engagement, and the qualitative part consists of a summary of insightful comments from the optional open question.

7.1.1 Reflection

The average TSRI [11] score across all conditions was M = 38.36, SD = 11.98, with the lowest score of 9 being recorded for the *MediumC* condition and the highest score of 63 being recorded for the *HighC* condition. Using R for the analysis (*https://www.r-project.org/*), I conducted a two-way ANOVA to determine the effect of the Abstraction level and Goal Completion on the TSRI score, where the trait reflection was controlled. Figure 7.1 shows box plots for the three levels of abstraction.

There was a significant effect of the Abstraction level on the TSRI scores, $F(2, 215) = 4.483$, $p < 0.05$, but no significant effect of Goal Completion was found. A significant effect of the trait reflection score, the covariant was found, $F(1, 215) = 7.738$, $p < 0.01$. Post-hoc tests using Turkey HSD revealed that there was a significant difference between the pairs *Low-Medium* and *Low-High* at $p < 0.05$, but no significant difference between the pair *Medium-High*. The perceived lack of difference in terms of reflection between Medium and High can also be seen in Table 7.1, with the two levels' mean and standard deviation being notably close.

Level	M	SD
Low	41.62	9.48
Medium	36.47	12.78
High	36.91	12.88

Table 7.1: The TSRI score mean and standard deviation for the three levels of abstraction.

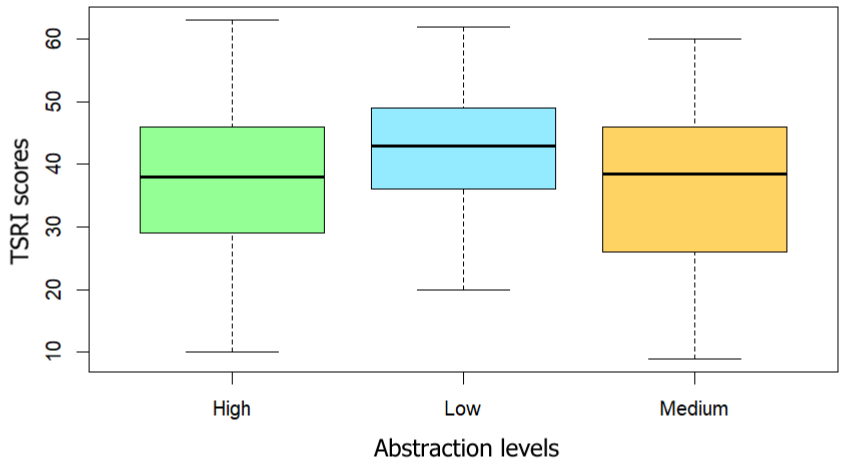


Figure 7.1: TSRI scores for the three levels of abstraction, where *Low* scored significantly higher than the other two levels.

I also analysed the subscales of the TSRI; Insight, Exploration, and Comparison. A significant effect of the Abstraction level was shown on the Insight and Exploration subscores, $F(2, 215) = 3.297$, $p < 0.05$, and $F(2, 215) = 5.702$, $p < 0.005$ respectively. No significant effect of the abstraction level was found on the Comparison subscale. However, with a p-value of 0.0554, the interaction effect between the abstraction level and the goal completion was close to being significant for this subscale. No significant effect of Goal Completion was found on any of the subscales. Table 7.2 shows the mean values of the subscale scores.

Level	Insight M	Exploration M	Comparison M
Low	13.51	15.37	12.74
Medium	11.82	13.12	11.52
High	12.13	13.34	11.43

Table 7.2: The TSRI subscale score means for the three levels of abstraction.

7.1.2 Engagement

The average UES-SF [79] score across all conditions was M = 37.86, SD = 8.44, with the lowest score of 15 being recorded for the *LowNC* condition and the highest score of 56 being recorded for the *HighNC* condition. A two-way ANOVA was used to determine the effect of the Abstraction level and Goal Completion on the UES-SF score. No significant effect of the Abstraction level, Goal Completion, nor any interaction effect was found.

Condition	**M**	**SD**
LowNC	35.81	9.31
LowC	39.57	7.65
MediumNC	38.47	7.74
MediumC	37.13	8.11
HighNC	36.81	9.16
HighC	39.36	8.44

Table 7.3: The UES-SF score mean and standard deviation for the six conditions.

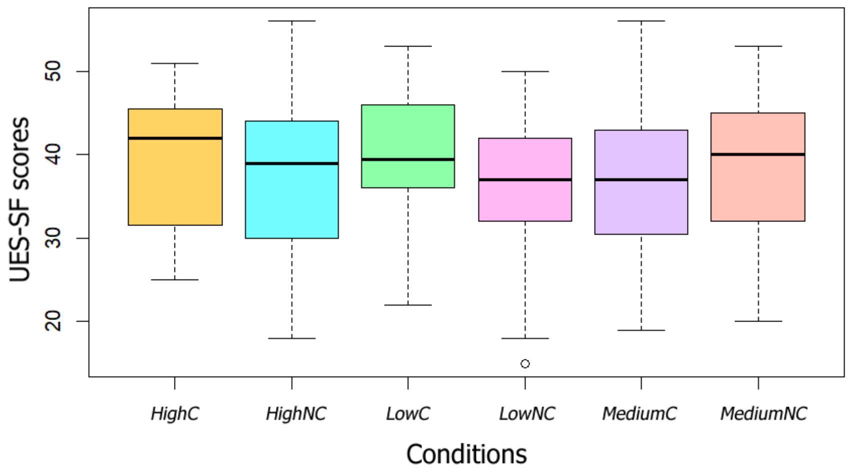


Figure 7.2: UES-SF scores for the six conditions, where no significant effect was found.

7.1.3 Open question

For the optional open question, the majority of participants chose not to leave any response. Out of the responses, most were on the condition with *Medium* or *High* abstraction. The responses left on the conditions with *Low* abstraction were mostly positive; *"Everything was great."* (P52, *LowNC*), *"Easy to understand"* (P157, *LowNC*), and *"It's clear and concise"* (P200, *LowC*).

A few responses were left for the conditions with *Medium* and *High* abstraction, voicing different concerns. Some participants found it difficult to interpret the progress being visualised: *"It didn't make it easy to know whether the steps undertaken were enough - it didn't offer any detail and I wouldn't find this overly useful."* (P32, *HighNC*), and *"Makes it difficult to see how much I have left to do, because all the pieces of the butterfly are not the same size."* (P49, *MediumNC*). Some also thought that the condition with *High* abstraction, in particular, was not specific enough: *"the blurred outline doesnt help at all. not specific enough"* (P134, *HighNC*), *"I felt it didn't really give any specific information."* (P178, *HighNC*), and *"It's too simple, too few information."* (P204, *HighC*).

Other concerns were the usefulness of the designs when comparing days; *"too abstract a way to track it and compare against other days etc."* (P40, *HighNC*). Additionally, another participant expressed interest but believed the design could be revised; *"I thought having a visualization was interesting though I believe there are versions that may fit better than a butterfly."* (P11, *MediumC*).

Some participants brought up the lack of numbers to accompany the visualisations; *"It has no actual data, with numbers, but numbers can also be boring with time, its good visualization but with numbers it would be better."* (P110, *MediumNC*), *"I think numbers help establish a sense of accomplishment. [...]"* (P175, *MediumNC*), *"The visualisation was too abstract. I prefer numbers."* (P167, LowC), and *"Added a percentage along with the visualization would be good."* (P122, *MediumNC*).

As for ways to improve the design, one participant brought up the use of colour customisation, which could make the visualisations more personal; *"I think allowing personalization of colors would be great! "* (P173, *MediumNC*).

7.2 Study II Results

The results from Study II are divided into a quantitative and a qualitative part. The quantitative part consists of the usability scores measured with the SUS [14], and the qualitative part consists of the themes emerging from analysing the interview data. Table 7.4 shows the demographic data of the participants.

Participant	Gender	Age	SUS score
1	Female	23	85
2	Female	43	90
3	Male	58	87.5
4	Female	56	57.5
5	Female	26	80
6	Female	21	87.5
7	Female	54	97.5
8	Male	22	60
9	Female	30	90

Table 7.4: Demographic data of the participants in study II, together with their SUS score.

7.2.1 Usability

Based on the answers from the questionnaire, each participant's SUS score was calculated following the directions from the author [14]. The average score was 81.66 (SD = 13.80), and each score can be seen in Table 7.4. Since a score above 68 is considered above average and a score above 75 is considered good, the usability of FlyFit should not have affected the experience negatively overall.

7.2.2 Interviews

The interviews were transcribed and then analysed using Thematic Analysis, as described by Blandford et al [3], with open coding. This analysing method has been used in other works on personal informatics [12, 74]. The analysis was done using the Atlas analysing software *https://atlasti.com/* and the initial coding was done through the software's automatic AI coding. The initial codes were revised and new codes were added through an iterative process. A total of 58 codes were generated and then grouped together to create three themes. The identified themes, Aesthetic representation, Enticement, and Comprehension, are described below together with excerpts from the interview data.

Aesthetic representation: When shown the visualisations, the majority of participants preferred *Medium* abstraction and expressed strong positive feelings towards it. Fun, pretty, interesting, and original was a few words used to describe it, including:

> *Beautiful. I love it [Medium]. It's the best one.* (P1)

> *It's [Medium] fun, it's new. It's pretty.* (P5)

Most participants liked *Low* abstraction but were neutral towards *High* abstraction. For most interview questions, participants answered them based on, and compared, *Medium* and *Low* abstraction, expressing disinterest in *High* abstraction altogether:

> *I don't feel much from the second butterfly.* (P7)

Participants were encouraged to think about the shape of *Medium* and *High* abstraction and what it represents. The participants clearly recognised it as a butterfly and were generally very positive toward the shape, founding it fitting. Because of how butterfly wings look in nature, it made sense to have the sections like the ones present in *Medium* abstraction. Additionally, while one participant mentioned seeing no connection between the butterfly and activity, a number of the participants could. Participants indicated that the butterfly represented transformation, freedom, and rapid movement, finding a butterfly to be pretty active. Reflecting on other options, the only other suggestion was a tree and mentions of things that would be inappropriate to use, such as a spider. Inanimate objects were also found to be unfitting:

> *No, I think a butterfly is quite fitting and a good choice, a dog or cat would be kinda random and butterflies kinda represent transformation*

> *and endurance or something along the likes of that and I associate keeping track of your steps with that since it's most likely to make sure you work out or move enough. Also, butterflies are pretty enough to kinda keep your attention, an inanimate object would also just be kinda random.* (P6)

Most participants appreciated the colour choice and expressed their positive feelings for blue. This, together with the use of a gradient, made some participants more appreciative of the design of *Low* as it appeared more aesthetically pleasing than regular bar charts tend to be. In relation to colours, some participants also mentioned the possibility of customisation, in order to make the visualisations appealing to a wider group of people:

> *Love it, love blue, love gradients, love the contrast so the colours are perfect in my opinion, maybe it'd be cool for the option to change the colour in case someone's fav colour is green or something. Very pretty.* (P6)

Just like with the shape of the design, the participants reflected on the current colour and what it could symbolise. Blue was thought to represent sadness and was perceived as cold by some, while others thought blue could symbolise calmness and comfort. One participant mentioned relating blue to hospitals and thus seeing the colour as healthy, while another thought blue was a happy colour because of the love the participant had for said colour. Even though blue wasn't seen as the most motivating colour or as having a connection to activity, it was preferred over having a colour that could indicate pressure or be stressful:

> *Blue is a calming colour. Red and orange would have been worse. Even though it is for physical activity, calm colours are better than stressful ones.* (P3)

Enticement: Having a clear preference for *Medium* abstraction and its design, participants saw the butterfly as a motivation to use the application and be active. Participants referred to it as "filling in" the butterfly or completing a puzzle, which increased interest and created an incentive to interact with the visualisation:

> *I think it could be motivating. I mean, uh, like putting a puzzle together. You're sort of getting closer to this goal in a sort of slightly vague way, which is kind of nice.* (P5)

The smaller butterflies which represented progress past the goal in *Medium* and *High* were clear to all participants. They were perceived as "rewards" for completing a goal and obtaining them was referred to as an "achievement" or "success". Having this aspect included in the design was found to be motivating for participants to reach, and even pass their goal:

> *I think I've done very, very well. Because especially like on this one [Low], yes, you can see the extra stuff like above, but that's basic. But when you see the butterfly with like extra butterflies. That's like oh. I*

> *did. I did better than I was supposed to. [...] It gives you a bit of an incentive to like go above the goal. And to do more.* (P1)

Comprehension: *High* abstraction was found to be difficult to interpret. On the other hand, all participants thought *Low* abstraction was clear and easy to use, having little trouble interpreting their results from it. Some participants felt confident interpreting their progress using *Medium* abstraction as well:

> *I can clearly see in the bar chart that this [the goal line] is the goal, and in the butterfly, I can see that the goal is to fill in the butterfly. The section in the butterfly is a percentual part of the process I believe. The differences in sizes have no meaning, I think, and each shape is an equal part of the whole thing. For me, this makes my current progress clear.* (P3)

One aspect affecting participants' comprehension of *Medium* was the use of the colour blue with different levels of brightness, not as a gradient but as several single values. The brightest blue appeared as white, which made some participants question if this meant the section was not completed or if they had not reached their goal yet:

> *It confused me a little bit in the beginning, cause I was like, OK, so I was thinking maybe that it was that the light colours were worth more or less or something, and together they reached a value in some way, but I assume now that they were just static colours that then were filled in random or by a determined span of some sort. So I mean, I guess they were confusing, but it's not a problem if you just get used to it.* (P5)

Another aspect, which was brought up by next to all participants, was the need for numbers. The desire was to have either the exact number or the percentage displayed, but it was not needed to be available at all times. Nevertheless, there was a strong desire that numbers should be available in some way if the user wants to know his exact progress. This desire was strongest for *Low* abstraction because of its design as a bar chart and thus feeling "natural" to have numbers. For *Medium* and *High*, some participants thought the current relative indication was enough, while some considered numbers to be necessary there as well:

> *As in the diagram with the bars is always very easy to read, but I missed in the bar, numbers. I would like to have seen a scale on the bar. I would like it to have a number showing how many percentages are finished in the other ones [Medium/High].* (P4)

As mentioned in related work, the benefits of having multiple visualisation options were brought up frequently in the interviews as well. Participants wished to have the option to choose since many believed *Medium* or *High* abstraction didn't work alone because of the difficulties in making an accurate interpretation. When the notion was reflected on more deeply, participants thought the best idea was to have only one visualisation displayed at once and that it could be switched out in settings for example:

> *I definitely think it could be nice if you were allowed to choose which one of the visualisations you want to have as like, your standard, but then also being able to change that or like go and look at the other ones. I mean seeing the data in this way in the butterfly is a bit different than seeing it in this way [Low]. Like if you have this one [Low] you can more clearly, in a way I guess see it climbing upwards. Uh, but this is [Medium] definitely a more fun way of visualising it. And I feel like there's more like an incentive to keep walking, to keep filling in the squares. So I think having a choice is definitely the best way to go about it.* (P1)

8

Discussion

In this thesis, I explored to what extent abstractness has an effect on relevant fitness-tracking aspects such as reflection and engagement. This was done through an online survey followed by interviews. This section discusses the findings and how they can be used in future research on personal informatics and the use of personal informatics tools. Motivations for choices as well as limitations are also included.

8.1 The Process

An aspect of the process that could be commented on is the fact that Study II was started before Study I was completely finished. Although the direction of Study II relied on the results from Study I, this decision was made due to the time constraint of the project. However, Study II wasn't started until the results from Study I indicated in which direction it would go in, following the recruitment of the first batch of participants. I thus believe that this arrangement of the process did not affect the outcome of the studies or the results.

The results from Study II were largely in line with the results from Study I, which validates the method used during this project as similar results were achieved through different study designs. Nevertheless, when deciding on the study design, many factors and options were taken into consideration. For example, instead of using Prolific to recruit participants for Study I, flyers could have been handed out on campus as there are often doubts about the validity of participants recruited online. The kind of manual work that comes with handing out flyers has other limitations though, mainly the reduced number of participants that would likely be recruited. A too-small sample size is a factor that decreases external validity. Additionally, about half of the participants recruited for Study II were people from campus, and since the results were similar to Study I, the validity of the Prolific recruited participants is strengthened.

For Study I, a pilot study was run after the study design had been decided on in order to verify that the design was acceptable before the survey was released to a large number of participants. A pilot study was also run after the context had been added to the visualisations and since no changes were made following that, the results from the second pilot study were included in the overall results. Running pilot studies is a way to reduce systematic errors, also called biases [56], which are error that offsets results in one direction and reduces the reliability of an experiment. Specifically bias caused by the experimental procedure can be reduced following the run of a pilot study.

Although reasons for why interviews and a survey were used have already been given in the chapters dedicated to each study, I will motivate those decisions a bit further. First, one strength of using a survey was that bias caused by an experimenter's behaviour [56] could be removed, since there was no direct interaction between the experimenter and the participants. Additionally, using a survey and interviews allowed for the collection of individual opinions. The solution for the studied problem requires an adaption, and individuals have different opinions on what the best fit is. Group opinions are not the answers relevant to this project, which further explains the decision to not use focus groups.

One design motivation for the application created during this project was the use of a daily goal in order to be flexible for users' different opportunities on different days. This problem can also be solved by having primary and secondary goals, which have been shown to be appealing [71]. Although this aspect might not have been a crucial factor for the turnout of study II, it is still worth mentioning that this design issue can be handled in several ways. The chosen way was simply used because of the better fit it had for the study and its purpose.

8.2 Findings

The results can be summarised as *Low* abstraction being the clearest and easiest to use, while *Medium* abstraction was considered most motivating and appealing. *High* abstraction wasn't particularly interesting altogether. This is similar to what has been found in previous research, which was also mentioned in Chapter 2, *Related Work*, where previous studies found charts to be better for readability and finding specific information, while abstract visualisations were more appealing and sparked an emotional response [34, 96].

8.2.1 Reflection

The results from Study I showed that the Abstraction level had a statistically significant effect on the TSRI score, which was used to measure reflection. The visualisation with *Low* abstraction performed significantly better than *Medium* and *High*, which means that *Low* abstraction was better at supporting users in reflecting on their progress and data. This was also shown in Study II, where all participants expressed finding *Low* to be the most understandable and easiest to interpret. The factors that made participants in Study II express their overall preference for *Medium* abstraction included motivation, appeal, and novelty. None of these aspects was measured during Study I, so whether *Medium* actually performs significantly better on these aspects is still unknown.

Despite past work indicating that people reflect to different extents depending on if they have completed their goal or not [40], the results showed that Goal Completion had no effect on the TSRI score. This can also be considered a positive result since the visualisations were designed in a way that didn't appear to affect the user in a negative way when the goal had not been completed. In addition, one of the behaviour change principles stated that technologies should reward desired behaviour and not punish undesired behaviour [24]. These results validate that the design indeed can be considered Positive, together with how participants referred to the small butterflies (showing the goal being completed) as rewards or achievements in Study II.

While *Low* abstraction had a higher mean score for all subscales of the TSRI, it was only significant for the Insight and Exploration subscales. *Low* was thus better at providing insightful information to the user as well as easier and more enjoyable to use for exploring the data, while not significantly better when it comes to using the visualisation to compare with others. This could be due to the fact that the visualisations only visualised the user's own (hypothetical) data together with the goal, which inherently is a representation of the user's own "battle" to fulfil the daily goal he has set out. This might be unfitting to compare with others. The data in question is also considered personal, which can be a reason for the low eagerness to share it with others. This hypothesis was partly supported during Study II. When participants were asked to send their daily step count, if they had done badly that day, according to themselves, many were quick to give a "heads-up" and explain why that was the case. Some also suggested using the data from another day where they had performed better instead. This stresses the need for abstraction, and other methods, to keep the data private when used or displayed in a public setting.

8.2.2 Engagement

There was no significant effect of Abstraction or Goal Completion on the engagement score, but there are still some interesting aspects that can be seen from studying the results. For instance, for the conditions with *Low* and *High* abstraction, the engagement scores were generally higher when the goal had been completed compared to when it had not. For *Medium*, the case was the opposite.

It is also interesting to note that *High* abstraction had the largest standard deviation. It appears that the higher the Abstraction level of the visualisation is, the greater the variation is between how people find it engaging. My interpretation of this is that abstractness, and its usefulness, are perceived differently by different people and could possibly be affected by certain personality traits, just like a person's reflectiveness is affected by trait reflection. This is somewhat supported by a comment left by a participant in study I, finding the *Low* abstraction visualisation to be "too abstract".

8.2.3 Design issues

The importance of including numbers in visualisations was brought up by several participants in both Study I and II. These comments were made on both *Low* and *Medium* abstraction. No such comment was specifically mentioned on *High* abstraction, but given the results overall, it is possible this would be desired for *High* as well. Cooper [25] has brought up the fact that, while visualisations are highly effective and should be used, they should not replace raw data (i.e. numbers) but rather be used together. Tracking is also often associated with exact numbers, which can make users expect feedback to be presented in a certain way [41]. However, since the purpose of the studies was to compare the visualisations and people's abilities to use the visualisations to collect insights on their progress, all other elements were excluded. On the other hand, the results might have turned out differently if numbers had been included, and it is clear that it is necessary to have numbers available in some way for an official application to be successful.

A notable comment that was made by a few of the participants in Study II was that the colours in *Medium* were confusing. Unlike *Low* and *High*, *Medium* does not have a gradient and instead uses single colour values, because of the sections. This, together with the decision to have the sections filled in at random seems to have made participants lose the connection between deeper colours and being closer to the goal, a connection one participant pointed out. The loss of connection is naturally understandable given the current design, and this is something that would need to be addressed if the design is going to be studied further in the future. Two ways to solve it are to either have *Medium* abstraction grow from the middle like *High* does or to have the shape be a gradient like *High* but still have the sections filled in at random. Figure 8.1 shows what the second suggestion would look like.

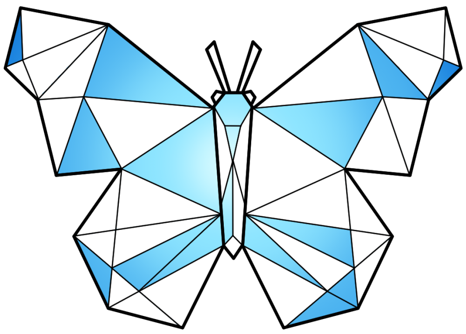


Figure 8.1: An alternative version of the *Medium*, based on responses from Study II.

Although not mentioned as often, there were also comments about if the different section sizes had any meaning. This was not the case, but it is understandable that participants questioned it. Nevertheless, the design was also supposed to be abstract, so every part of the design will not necessarily be clear to every user. However, it is still important that a visualisation can be interpreted by the consumer fairly easily or it will not be particularly useful. Insights can, after all, not be derived from data unless it is interpretable, understandable, and accessible [48]. Evidently, the studies showed that there are a number of issues with the design created during the work of this thesis and that there is room to improve the designs to make them more effective.

8.2.4 Privacy

The main motivation for exploring abstract visualisation in the first place was the need for privacy, something not always discussed within personal informatics despite its importance [31]. Somewhere along the work of this thesis, the privacy factor was somewhat lost. The vignette in Study I included the visualisations being viewed in public, but the scales used to measure the visualisations don't focus on the public aspect. So while *Low* abstraction was significantly better at supporting reflection, the other designs still have the potential for other aspects of the tracking experience. Interestingly enough, the only subscale *Low* abstraction wasn't significantly better on was Comparison, the scale closest to tackling the public aspect.

Evidently, people still had strong feelings towards the more abstract designs, in particular the *Medium* one considering the responses from Study II. In addition, *Medium* and *High* abstraction certainly are more private than the *Low* abstraction bar chart, since it is not obvious to outsiders what they represent. As I said, the privacy aspect was partly lost during the process and more focus could have been put into it during the interviews to explore that aspect more. The interviews showed, after all, that participants saw benefits to having several visualisation options. Using both a visualisation that supports reflection and is easy to accurately interpret, such as the bar chart, and using an abstract visualisation that is fun, motivating, and private, could potentially lead to the most comfort for the user together with the enticement to track long term. This is supported by previous work suggesting that engagement and enjoyment can be increased over time following the use of both abstract art and charts [34, 96].

Dividing user interactions with their tracking devices into glances and engage sessions was mentioned in Chapter 2, *Related Work*. Thinking in those terms and considering past research [34, 96] as well as the research done in this thesis, the bar chart could possibly be the best fit for engage sessions while an abstract design could be the best fit for glances. While in public, the most prominent type of interaction is likely glances. It is thus not necessary in those situations to have a visualisation that allows the user to reflect and interpret an exact progress status (i.e. a bar chart for example). Glanceable feedback has after all a strong effect on an individual's behaviours and as mentioned by one of the participants in Study II, just knowing that the user needs to walk *more* might be enough when out in public. So, there is partly evidence that abstractness contributes to privacy.

8.2.5 Ease of Use VS Attractiveness

An important finding was that *Low* abstraction was easier to use while *Medium* abstraction was considered more attractive because of its appeal and motivational properties. A simple way to describe the two terms in this context would be that ease of use is how easily a user can obtain insights, understand her progress, and get the answers she seeks, while attractiveness would describe a visualisation being pleasant and appealing to look at and interact with. When something is easy to use, it can become attractive, but generally not the other way around. Designing a visualisation that has both aspects or finding a good balance between the two has proven to be difficult during the work of this thesis.

It is possible that the higher attractiveness of *Medium* abstraction was due to novelty though, as the design was something new while *Low* abstraction was of a bar chart that people commonly are presented with. To properly establish the attractiveness of *Medium* abstraction, it would need to be studied during a more extended period of time.

8.3 Social Impact

The two studies, Study I and Study II, used the visualisations designed during this project, called *Low*. *Medium*, and *High* abstraction. These specific designs representing different levels of abstraction emerged from the ambition to explore concrete and abstract visualisation techniques. To relate my results to the initial area of interest, my results indicate that concrete visualisations are better for supporting users (ease of use), while abstract visualisations appear to be more promising regarding attractiveness.

Since these studies were conducted using step count as the data, a commonly tracked area within fitness tracking, it is likely possible to generalize the findings to other areas within fitness tracking as well. It is unknown if it is possible to generalize the results to other areas of personal informatics though, such as finance and location for example.

The motivations for exploring fitness tracking, in particular, are health and well-being. Changing behaviours and obtaining a healthier lifestyle are the main motivations for tracking physical activity, but the many existing challenges hamper many trackers' progress. Developing better visualisations is just one of the many aspects that must be explored to reduce those challenges. My contribution to the field of personal informatics is that we now know a little more about different types of visualisations fitting for displaying fitness tracking data. The goal of exploring personal informatics is not to design the very best tool but to design tools that help people in the process of improving their health and well-being.

8.3.1 Ethical Issues

A number of ethical issues were taken into account during the work of this thesis. Since the thesis' focus was to study visualisations with the use of a mobile interface, the ethical issues were mainly those related to mobile interfaces. A prominent issue with mobile interfaces is accessibility. For example, a large part of the population has low literacy, and those users need certain considerations when designing a mobile interface [17]. Considerations should also be taken for novice users [69], older people [29], and people with disabilities [7], which are just a few examples.

Another ethical concern is the use and collection of personal data, which is the very core of self-tracking. Personal data can be sensitive which leads to issues about privacy. A mapping of personal informatics literature showed that only about a third of the publications reviewed explicitly mentioned privacy or ethical concerns [31]. Some system actions to mitigate privacy concerns were suggested in these publications, such as encrypting data, limiting the captured data types, and keeping data abstract.

While self-tracking can be helpful and support people in changing their behaviour for the better, it can also be negative for some people's mental and physical well-being. For example, while a person can become aware of her stress levels when presented with her stress data, seeing it displayed can also intensify her stress [50]. Because of this, it is important how and when feedback is displayed, and the way it is represented matters. Self-tracking can contribute to a problem if the person has obsessive habits, and it is therefore not always a good management strategy to closely monitor one's behaviour.

Using negative framing techniques (for example highlighting unmet goals rather than achieved goals in charts) when presenting tracking data can evoke strong negative emotions in users, such as frustration, annoyance, and discouragement [63]. Negative feedback can also make users feel stressed and pressured [74]. Instead, data should be displayed in a way that can empower people to grow and improve their ability to cope with potentially unmet goals which often is perceived as a failure. If there is a compelling reason for displaying negative data, it should be done in a way that doesn't discourage or demotivate the user [50].

Considering the scope of this thesis, it was not possible to focus on all of these aspects. Ultimately, the focus was on privacy and how to keep data abstract by exploring abstract visualisation techniques. Accessibility was also focused on to some degree by using Material Design, which has accessibility standards built into code and design, such as scalability across different screen sizes. Lastly, avoiding negative framing techniques and instead keeping the visualisations neutral and/or positive was also an important aspect of the designs developed.

8.4 Limitations

A limitation of Study I was the use of a vignette. A belief does not equal action. Just because the results show that participants believed *Low* abstraction offered more opportunities for reflection doesn't necessarily mean that people will actually reflect more following the use of that visualisation when used "in the wild". However, previous work on the vignette study method has shown that it can generate similar results to what real-world experience would [44] which is why the method can still be highly effective. Another limitation was the use of hypothetical data instead of participants' own data, which decreased the ecological validity [12]. A number of confounding factors emerge from exploring personal data though, which wouldn't be possible to reliably measure. This is the motivation for the study design being the way it is. The use of personal data was instead used in Study II to allow for exploration where the setup and data collection were different.

There were also limitations to Study II. Since this study used a designed interface, the feedback received from the participants was influenced by it. While the design was very simplistic and followed many guidelines and principles of designing an interface, any design will still have an inherent bias [12]. To increase the validity of results, different interface designs should be explored and studied. Additionally, there was also an uneven gender distribution. For Study I, this was controlled through the recruitment process, but due to the convenience sample used in Study II, this aspect wasn't controlled to the same extent. This uneven distribution has occurred in other personal informatics studies as well, such as having a 20% and 80% distribution [2], or even 100% of one gender [41]. While not appearing to be detrimental considering previous work, a better distribution would have been desirable and also better for the result. The convenience sample recruitment method could also have caused social desirability bias to some extent, with the participants wanting to give positive feedback because of their relationship with the researcher. However, 1/3 of the participants from Study II were not familiar with the researcher prior to the interviews, so this bias was likely limited.

One criterion for the participants was that they should have limited experience with tracking due to some aspects of the visualisation designs (such as single-day data). This was not measured in Study II since the convenience sampling method was used, so the tracking experience could be regulated. For Study I, the reported mean value for the tracking experience was 3.45, which was a bit higher than what was originally desired. Possible bias because of the participants' tracking experience should thus be recognised. In addition, possible bias because of the participants' experience with analysing visualisations, and in particular, abstract visualisation, was acknowledged but ultimately not measured in either of the studies.

Another important limitation to note is that the project used a population from a Western background for both studies. Because of this, the applicability of the work is limited since cultural differences affect how people view different things. This is however in line with many other published works within personal informatics, but acknowledging the limitation is still necessary. The project was also conducted from a Western perspective which together with my positionality results in potential biases and influences the work as a whole.

8.5 Future Work

I believe that an important subject to explore further is the use of multiple visualisations where the user can decide which to use at a given moment. Comparing the usage of a bar chart and an abstract design, to what extent the visualisations are used and in what situations the user decides to use which. Exploring if indeed the bar chart is used for engage sessions while an abstract design is used for glances, or if one of them is preferred for both types of interaction. This suggestion is motivated by how both my and previous work has shown the desire users have for being able to choose between visualisations [80]. Described by Shneiderman [90] as detail-on-demand, users generally prefer to get an overview with visualisations but want details to be available if they desire more information. Exploring this is important because of the challenges and design issues that still remain with tracking devices.

Another aspect to study further is the customisation of visualisations in tracking applications. Allowing for customisation was mentioned by participants in both Study I and II. Something as simple as changing the colour of a bar chart can make it go from being considered dull and boring to strongly appealing to someone. This is supported by how one participant in Study II mentioned how much more appealing the bar chart was compared to regular ones just because of the gradient used. Customisation makes something more personal, and by being allowed to personalise one's space, it becomes more likeable, familiar, and pleasant to be in [25]. It would be interesting if research could explore if allowing customisation helps increase engagement in the personal informatics context.

This thesis only explored abstractness using one abstract shape, in the form of a butterfly. Even though a difference appeared between the Abstraction levels for these designs, it cannot be guaranteed that the same difference will appear in other designs with different levels of abstractness. To strengthen the suggestion that abstractness does affect reflection and that abstract visualisations are superior to concrete visualisations in terms of attractiveness, more abstract designs should be explored. It is also of interest to explore more shapes to find out what the best type of representation could be. While one participant mentioned in Study II how an inanimate object would be "random", another person might feel like a "living being" would create pressure because of the feeling of needing to keep that living being happy.

9

Conclusion

This thesis explored abstractness in visualisations when used for fitness trackers and fitness tracking data. Three visualisations with different levels of abstraction were designed through an iterative and user-involved process, where two pre-studies were conducted to involve users. A simple mobile application called FlyFit was then designed and implemented. This application lets users set a daily goal and asses their progress using the different visualisations. The work was evaluated through two studies; an online survey and interviews.

I found that *Low* abstraction performed significantly better in terms of supporting reflection in users, which indicates that abstractness indeed has an effect on self-reflection. No significant effect of the Abstraction level was found on engagement, which suggests that abstractness does not affect engagement, but that other factors do instead. I also found that *Low* abstraction was the easiest to use and interpret, while *Medium* abstraction was considered the most appealing, motivating, and a design that could work as an incentive to continue tracking. This indicates that concrete visualisations could be more appropriate in terms of ease of use while abstract visualisations could be better in terms of attractiveness. Having the possibility to choose which visualisation to use amongst alternatives was considered important by participants.

My results imply that abstract designs can't completely replace standard visualisations such as bar charts, but that there are benefits to utilizing both visualisation techniques to best support users. A potential way forward is to further explore the use of multiple visualisation options to gain a better understanding of how users behave and react following this feature. I hope that this work will inspire more research into the abstraction of fitness tracking data and how it can be used to increase privacy while still supporting users.

Bibliography

[1] Herman Aguinis and Kyle J. Bradley. Best practice recommendations for designing and implementing experimental vignette methodology studies. *Organizational Research Methods*, 17(4):351–371, 2014.

[2] Fereshteh Amini, Khalad Hasan, Andrea Bunt, and Pourang Irani. Data representations for in-situ exploration of health and fitness data. In *Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare*, PervasiveHealth '17, page 163–172, New York, NY, USA, 2017. Association for Computing Machinery.

[3] Stephann Makri Ann Blandford, Dominic Furniss. Qualitative hci research: Going behind the scenes. *Synthesis Lectures on Human-Centered Informatics*, 2016.

[4] Christiane Atzmüller and Peter Steiner. Experimental vignette studies in survey research. *Methodology: European Journal of Research Methods for The Behavioral and Social Sciences*, 6:128–138, 01 2010.

[5] Leo C. Aukes, Jelle Geertsma, Janke Cohen-Schotanus, Rein P. Zwierstra, and Joris P.J. Slaets. The development of a scale to measure personal reflection in medical practice and education. *Medical Teacher*, 29(2-3):177–182, 2007. PMID: 17701630.

[6] Amid Ayobi, Tobias Sonne, Paul Marshall, and Anna L. Cox. Flexible and mindful self-tracking: Design implications from paper bullet journals. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, page 1–14, New York, NY, USA, 2018. Association for Computing Machinery.

[7] Mars Ballantyne, Archit Jha, Anna Jacobsen, J. Scott Hawker, and Yasmine N. El-Glaly. Study of accessibility guidelines of mobile applications. In *Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia*, MUM 2018, page 305–315, New York, NY, USA, 2018. Association for Computing Machinery.

[8] Aaron Bangor, Philip T. Kortum, and James T. Miller. An empirical evaluation of the system usability scale. *International Journal of Human–Computer*

Interaction, 24(6):574–594, 2008.

[9] Mark Baskinger. Pencils before pixels: A primer in hand-generated sketching. *Interactions*, 2008.

[10] Frank Bentley, Konrad Tollmar, Peter Stephenson, Laura Levy, Brian Jones, Scott Robertson, Ed Price, Richard Catrambone, and Jeff Wilson. Health mashups: Presenting statistical patterns between wellbeing data and context in natural language to promote behavior change. *ACM Trans. Comput.-Hum. Interact.*, 20(5), nov 2013.

[11] Marit Bentvelzen, Jasmin Niess, Mikołaj P Woźniak, and Paweł W Woźniak. The development and validation of the technology-supported reflection inventory. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–8, 2021.

[12] Marit Bentvelzen, Jasmin Niess, and Paweł W. Woźniak. Designing reflective derived metrics for fitness trackers. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 6(4), jan 2023.

[13] David Benyon. *Designing user experience.* Pearson UK, 2019.

[14] John Brooke. Sus: A quick and dirty usability scale. *Usability Eval. Ind.*, 189, 11 1995.

[15] Hanington Bruce and Bella Martin. *Universal Methods of design.* Quarto Publishing Group USA, 2019.

[16] Bill Buxton. *Sketching user experiences: getting the design right and the right design.* Morgan kaufmann, 2010.

[17] Beenish M. Chaudry, Kay H. Connelly, Katie A. Siek, and Janet L. Welch. Mobile interface design for low-literacy populations. In *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*, IHI '12, page 91–100, New York, NY, USA, 2012. Association for Computing Machinery.

[18] Eun Kyoung Choe, Bongshin Lee, and m.c. schraefel. Characterizing visualization insights from quantified selfers' personal data presentations. *IEEE Computer Graphics and Applications*, 35(4):28–37, 2015.

[19] Eun Kyoung Choe, Bongshin Lee, Haining Zhu, Nathalie Henry Riche, and Dominikus Baur. Understanding self-reflection: How people reflect on personal data through visual data exploration. In *Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare*, PervasiveHealth '17, page 173–182, New York, NY, USA, 2017. Association for Computing Machinery.

[20] Eun Kyoung Choe, Nicole B. Lee, Bongshin Lee, Wanda Pratt, and Julie A. Kientz. Understanding quantified-selfers' practices in collecting and exploring personal data. In *Proceedings of the SIGCHI Conference on Human Factors*

in Computing Systems, CHI ’14, page 1143–1152, New York, NY, USA, 2014. Association for Computing Machinery.

[21] Sunny Consolvo, Predrag Klasnja, David W. McDonald, Daniel Avrahami, Jon Froehlich, Louis LeGrand, Ryan Libby, Keith Mosher, and James A. Landay. Flowers or a robot army? encouraging awareness amp; activity with personal, mobile displays. In *Proceedings of the 10th International Conference on Ubiquitous Computing*, UbiComp ’08, page 54–63, New York, NY, USA, 2008. Association for Computing Machinery.

[22] Sunny Consolvo, Predrag Klasnja, David W. McDonald, and James A. Landay. Goal-setting considerations for persuasive technologies that encourage physical activity. In *Proceedings of the 4th International Conference on Persuasive Technology*, Persuasive ’09, New York, NY, USA, 2009. Association for Computing Machinery.

[23] Sunny Consolvo, Predrag Klasnja, David W. McDonald, and James A. Landay. Designing for healthy lifestyles: Design considerations for mobile technologies to encourage consumer health and wellness. *Found. Trends Hum.-Comput. Interact.*, 6(3–4):167–315, apr 2014.

[24] Sunny Consolvo, David W. McDonald, and James A. Landay. Theory-driven design strategies for technologies that support behavior change in everyday life. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI ’09, page 405–414, New York, NY, USA, 2009. Association for Computing Machinery.

[25] Alan Cooper, Robert Reimann, David Cronin, and Christopher Noessel. *About face: the essentials of interaction design*. John Wiley & Sons, 2014.

[26] D Council. Framework for innovation: Design council’s evolved double diamond, Jun 2022. `https://www.designcouncil.org.uk/our-work/skills-learning/tools-frameworks/framework-for-innovation-design-councils-evolved-double-diamond/`.

[27] Pranab Dash and Y. Charlie Hu. How much battery does dark mode save? an accurate oled display power profiler for modern smartphones. In *Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services*, MobiSys ’21, page 323–335, New York, NY, USA, 2021. Association for Computing Machinery.

[28] Alan Dix, Janet Finlay, Gregory D Abowd, and Russell Beale. *Human-computer interaction*. Pearson Education, 2003.

[29] José-Manuel Díaz-Bossini and Lourdes Moreno. Accessibility to mobile interfaces for older people. *Procedia Computer Science*, 27:57–66, 2014. 5th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion, DSAI 2013.

[30] Daniel Epstein, Felicia Cordeiro, Elizabeth Bales, James Fogarty, and Sean Munson. Taming data complexity in lifelogs: Exploring visual cuts of personal informatics data. In *Proceedings of the 2014 Conference on Designing Interactive Systems*, DIS '14, page 667–676, New York, NY, USA, 2014. Association for Computing Machinery.

[31] Daniel A. Epstein, Clara Caldeira, Mayara Costa Figueiredo, Xi Lu, Lucas M. Silva, Lucretia Williams, Jong Ho Lee, Qingyang Li, Simran Ahuja, Qiuer Chen, Payam Dowlatyari, Craig Hilby, Sazeda Sultana, Elizabeth V. Eikey, and Yunan Chen. Mapping and taking stock of the personal informatics literature. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 4(4), dec 2020.

[32] Daniel A. Epstein, Jennifer H. Kang, Laura R. Pina, James Fogarty, and Sean A. Munson. Reconsidering the device in the drawer: Lapses as a design opportunity in personal informatics. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '16, page 829–840, New York, NY, USA, 2016. Association for Computing Machinery.

[33] Daniel A. Epstein, An Ping, James Fogarty, and Sean A. Munson. A lived informatics model of personal informatics. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '15, page 731–742, New York, NY, USA, 2015. Association for Computing Machinery.

[34] Chloe Fan, Jodi Forlizzi, and Anind K. Dey. A spark of activity: Exploring informative art as visualization for physical activity. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, UbiComp '12, page 81–84, New York, NY, USA, 2012. Association for Computing Machinery.

[35] Interaction Design Foundation. The 5 stages in the design thinking process. `https://www.interaction-design.org/literature/article/5-stages-in-the-design-thinking-process`.

[36] Interaction Design Foundation. What is human-computer interaction (hci)? `https://www.interaction-design.org/literature/topics/human-computer-interaction`.

[37] Interaction Design Foundation. What is user centered design? `https://www.interaction-design.org/literature/topics/user-centered-design`.

[38] Google. Google scholar, 2023. `https://scholar.google.com/`.

[39] Google. Material design, 2023. `https://m3.material.io/`.

[40] Rúben Gouveia, Evangelos Karapanos, and Marc Hassenzahl. How do we engage with activity trackers? a longitudinal study of habito. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiqui-*

tous Computing, UbiComp ’15, page 1305–1316, New York, NY, USA, 2015. Association for Computing Machinery.

[41] Rúben Gouveia, Fábio Pereira, Evangelos Karapanos, Sean A. Munson, and Marc Hassenzahl. Exploring the design space of glanceable feedback for physical activity trackers. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp ’16, page 144–155, New York, NY, USA, 2016. Association for Computing Machinery.

[42] Anthony Grant, John Franklin, and Peter Langford. The self-reflection and insight scale: A new measure of private self-consciousness. *Social Behavior and Personality: an international journal*, 30:821–835, 01 2002.

[43] Rebecca Gulotta, Jodi Forlizzi, Rayoung Yang, and Mark Wah Newman. Fostering engagement with personal informatics systems. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*, DIS ’16, page 286–300, New York, NY, USA, 2016. Association for Computing Machinery.

[44] Jens Hainmueller, Dominik Hangartner, and Teppei Yamamoto. Validating vignette and conjoint survey experiments against real-world behavior. *Proceedings of the National Academy of Sciences*, 112(8):2395–2400, 2015.

[45] Lars Hallnäs and Johan Redström. Slow technology - designing for reflection. *Personal and Ubiquitous Computing*, 5:201–212, 08 2001.

[46] S.F Herbes. The understanding of ambiguity as a design resource. Master’s thesis, Utrecht University, 2022.

[47] Kaspar Hornbæk. Some whys and hows of experiments in human–computer interaction.s. *Foundations and Trends in Human-Computer Interaction*, 5(4), 2013.

[48] Dandan Huang, Melanie Tory, Bon Adriel Aseniero, Lyn Bartram, Scott Bateman, Sheelagh Carpendale, Anthony Tang, and Robert Woodbury. Personal visualization and personal visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, 21(3):420–433, 2015.

[49] Apple Inc. Identify your iphone model, Jan 2023. `https://support.apple.com/en-us/HT201296`.

[50] Christina Kelley, Bongshin Lee, and Lauren Wilcox. Self-tracking for mental wellness: Understanding expert perspectives and student experiences. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI ’17, page 629–641, New York, NY, USA, 2017. Association for Computing Machinery.

[51] Kangsoo Kim, Austin Erickson, Alexis Lambert, Gerd Bruder, and Greg Welch. Effects of dark mode on visual fatigue and acuity in optical see-through head-mounted displays. In *Symposium on Spatial User Interaction*, SUI ’19, New York, NY, USA, 2019. Association for Computing Machinery.

[52] Young-Ho Kim, Jae Ho Jeon, Eun Kyoung Choe, Bongshin Lee, KwonHyun Kim, and Jinwook Seo. Timeaware: Leveraging framing effects to enhance personal productivity. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, page 272–283, New York, NY, USA, 2016. Association for Computing Machinery.

[53] Jurek Kirakowski. The software usability measurement inventory: background and usage. *Usability evaluation in industry*, pages 169–178, 1996.

[54] Jeffrey W Knopf. Doing a literature review. *PS: Political Science & Politics*, 39(1):127–132, 2006.

[55] Jon Kolko. *Thoughts on interaction design*. Morgan Kaufmann, 2010.

[56] Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. *Research methods in human-computer interaction*. Morgan Kaufmann, 2017.

[57] Ian Li, Anind Dey, and Jodi Forlizzi. A stage-based model of personal informatics systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, page 557–566, New York, NY, USA, 2010. Association for Computing Machinery.

[58] Ian Li, Anind K. Dey, and Jodi Forlizzi. Understanding my data, myself: Supporting self-reflection with ubicomp technologies. In *Proceedings of the 13th International Conference on Ubiquitous Computing*, UbiComp '11, page 405–414, New York, NY, USA, 2011. Association for Computing Machinery.

[59] Ian Li, Anind K. Dey, and Jodi Forlizzi. Using context to reveal factors that affect physical activity. *ACM Trans. Comput.-Hum. Interact.*, 19(1), may 2012.

[60] Zhuying Li, Yan Wang, Jacob Sheahan, Beisi Jiang, Stefan Greuter, and Florian 'Floyd' Mueller. Insideout: towards an understanding of designing playful experiences with imaging capsules. In Will Odom, Audrey Desjardins, and Marianne Graves Peterson, editors, *Proceedings of the 2020 ACM on Designing Interactive Systems Conference*, pages 601–613, United States of America, 2020. Association for Computing Machinery (ACM). Designing Interactive Systems 2020, DIS 2020 ; Conference date: 06-07-2020 Through 10-07-2020.

[61] Ying Lin. Smartphone market share worldwide [feb 2023 update]. `https://www.oberlo.com/statistics/smartphone-market-share`.

[62] Edwin Locke and Gary Latham. Building a practically useful theory of goal setting and task motivation: A 35year odyssey. *American Psychologist - AMER PSYCHOL*, 57:705–717, 09 2002.

[63] Meagan B. Loerakker. Design for personal informatics: Exploring the effect of personal data visualisations on cognitive processes. Bachelor's thesis, Utrecht University, 2021.

[64] Jonas Löwgren and Erik Stolterman. *Thoughtful interaction design: A design perspective on information technology.* Mit Press, 2004.

[65] Deborah Lupton. Self-tracking, health and medicine. *Health Sociology Review*, 26(1):1–5, 2017.

[66] I. S. MacKenzie. *Human-computer interaction: An empirical research perspective.* Elsevier Science Technology, 2012.

[67] I. Scott MacKenzie. *The Agile Developer's Handbook : Get More Value from Your Software Development: Get the Best Out of the Agile Methodology.* Packt Publishing, 2018.

[68] Keith Townsend Mark N.K. Saunders. Choosing participants. *The SAGE Handbook of Qualitative Business and Management Research Methods: History and Traditions*, 2019.

[69] Indrani Medhi, Somani Patnaik, Emma Brunskill, S.N. Nagasena Gautama, William Thies, and Kentaro Toyama. Designing mobile interfaces for novice and low-literacy users. *ACM Trans. Comput.-Hum. Interact.*, 18(1), may 2011.

[70] Sean Munson. Mindfulness, reflection, and persuasion in personal informatics. In *Extended Abstracts of CHI'12*, 2012.

[71] Sean A. Munson and Sunny Consolvo. Exploring goal-setting, rewards, self-monitoring, and sharing to motivate physical activity. In *2012 6th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops*, pages 25–32, 2012.

[72] Tamara Munzner. *Visualization analysis and design.* CRC press, 2014.

[73] G. Neff and D. Nafus. *Self-Tracking.* The MIT Press Essential Knowledge series. MIT Press, 2016.

[74] Jasmin Niess, Kristina Knaving, Alina Kolb, and Paweł W. Woźniak. Exploring fitness tracker visualisations to avoid rumination. In *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '20, New York, NY, USA, 2020. Association for Computing Machinery.

[75] Jasmin Niess and Paweł W. Woźniak. Supporting meaningful personal fitness: The tracker goal evolution model. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, page 1–12, New York, NY, USA, 2018. Association for Computing Machinery.

[76] Renee Noortman, Britta F. Schulte, Paul Marshall, Saskia Bakker, and Anna L. Cox. Hawkeye - deploying a design fiction probe. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, page 1–14, New York, NY, USA, 2019. Association for Computing Machinery.

[77] Claudia Núñez Pacheco and Lian Loke. Aesthetic resources for technology-mediated bodily self-reflection: The case of eloquent robes. In *Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures: The Future of Design*, OzCHI '14, page 1–10, New York, NY, USA, 2014. Association for Computing Machinery.

[78] Chalmers University of Technology. Chalmers library, 2023. `https://www.lib.chalmers.se/`.

[79] Heather L. O'Brien, Paul Cairns, and Mark Hall. A practical approach to measuring user engagement with the refined user engagement scale (ues) and new ues short form. *International Journal of Human-Computer Studies*, 112:28–39, 2018.

[80] Bernd Ploderer, Wolfgang Reitberger, Harri Oinas-Kukkonen, and Julia Gemert-Pijnen. Social interaction and reflection for behaviour change. *Personal Ubiquitous Comput.*, 18(7):1667–1676, oct 2014.

[81] Jenny Preece, Yvonne Rogers, Helen Sharp, David Benyon, Simon Holland, and Tom Carey. *Human-computer interaction*. Addison-Wesley Longman Ltd., 1994.

[82] Qualtrics. Qualtricsxm: Online survey software, 2023. `https://www.qualtrics.com/uk/`.

[83] Amon Rapp and Federica Cena. Personal informatics for everyday life: How users without prior self-tracking experience engage with personal data. *International Journal of Human-Computer Studies*, 94, 05 2016.

[84] Gretchen Reynolds. Do we really need to take 10,000 steps a day for our health?, Jul 2021.

[85] Yvonne Rogers, William R. Hazlewood, Paul Marshall, Nick Dalton, and Susanna Hertrich. Ambient influence: Can twinkly lights lure and abstract representations trigger behavioral change? In *Proceedings of the 12th ACM International Conference on Ubiquitous Computing*, UbiComp '10, page 261–270, New York, NY, USA, 2010. Association for Computing Machinery.

[86] John Rooksby, Mattias Rost, Alistair Morrison, and Matthew Chalmers. Personal tracking as lived informatics. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, page 1163–1172, New York, NY, USA, 2014. Association for Computing Machinery.

[87] Kim Sauvé, Saskia Bakker, Nicolai Marquardt, and Steven Houben. Loop: Exploring physicalization of activity tracking data. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*, NordiCHI '20, New York, NY, USA, 2020. Association for Computing Machinery.

[88] Wilmar B Schaufeli, Marisa Salanova, Vicente González-Romá, and Arnold B

Bakker. The measurement of engagement and burnout: A two sample confirmatory factor analytic approach. *Journal of Happiness studies*, 3:71–92, 2002.

[89] The Schedio. How can rounded buttons change the game for your design?, Jun 2020. `https://theschedio.com/rounded-buttons/`.

[90] B. Shneiderman. The eyes have it: a task by data type taxonomy for information visualizations. In *Proceedings 1996 IEEE Symposium on Visual Languages*, pages 336–343, 1996.

[91] Martin A. Siegel and Jordan Beck. Slow change interaction design. *Interactions*, 21(1):28–35, jan 2014.

[92] Gaurav Sinha, Rahul Shahi, and Mani Shankar. Human computer interaction. In *2010 3rd International Conference on Emerging Trends in Engineering and Technology*, pages 1–4. IEEE, 2010.

[93] Drew Skau and Robert Kosara. Arcs, angles, or areas: Individual data encodings in pie and donut charts. *Comput. Graph. Forum*, 35(3):121–130, jun 2016.

[94] Hannah Snyder. Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, 104:333–339, 2019.

[95] Danielle Albers Szafir. The good, the bad, and the biased: Five ways visualizations can mislead (and how to fix them). *interactions*, 25(4):26–33, 2018.

[96] Xin Tong, Diane Gromala, Lyn Bartram, Fateme, Rajabiyazdi, and Sheelagh Carpendale. Evaluating the effectiveness of three physical activity visualizations — how people perform vs. perceive. 10 2015.

[97] Paul Trapnell and Jennifer Campbell. Private self-consciousness and the five-factor model of personality: Distinguishing rumination from reflection. *Journal of personality and social psychology*, 76:284–304, 03 1999.

[98] Miriam Walker, Leila Takayama, and James A. Landay. High-fidelity or low-fidelity, paper or computer? choosing attributes when testing web prototypes. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 46(5):661–665, 2002.

[99] Martin Ward. A definition of abstraction. *Journal of Software Maintenance: Research and Practice*, 7(6):443–450, 1995.

[100] Colin Ware. *Information Visualization : Perception for Design.* Interactive Technologies Ser. Elsevier Science Technology, 2012.

[101] Anthony Wasserman. Software engineering issues for mobile application development. pages 397–400, 11 2010.

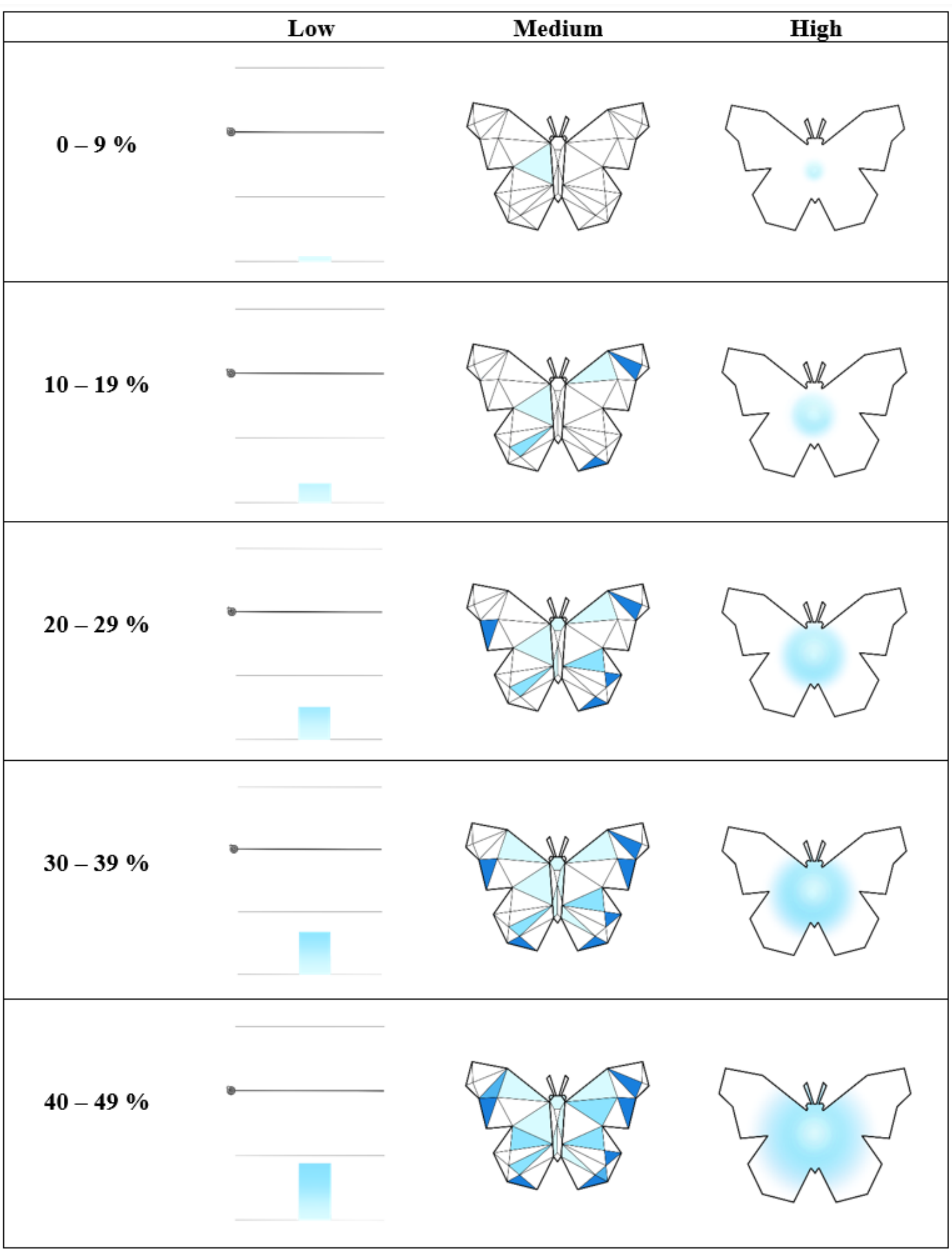
[102] Chauncey Wilson. *Brainstorming and beyond: a user-centered design method.*

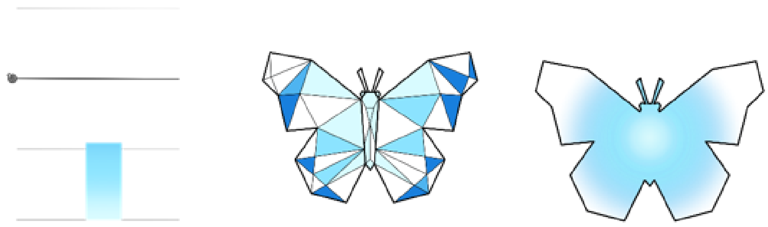
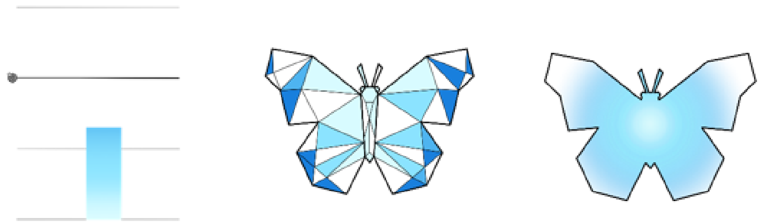
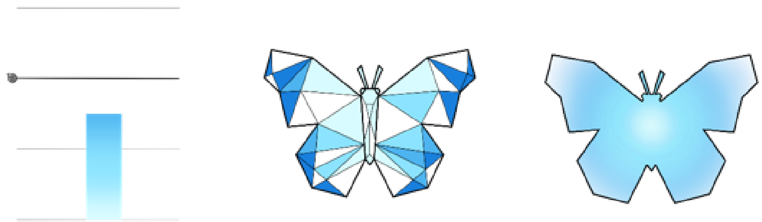
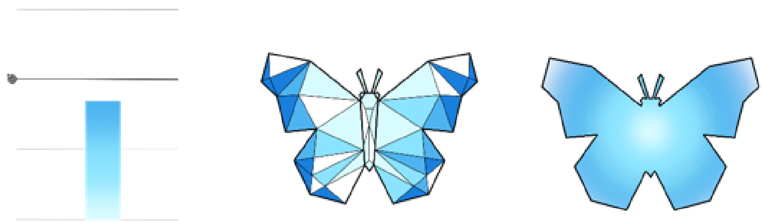
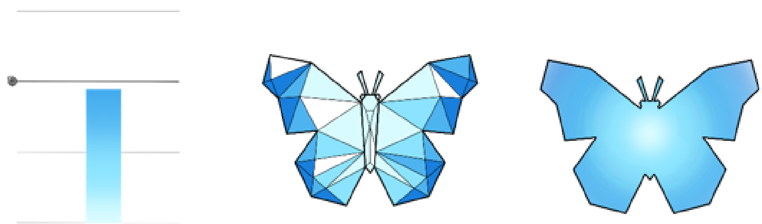
Morgan Kaufmann, 2013.

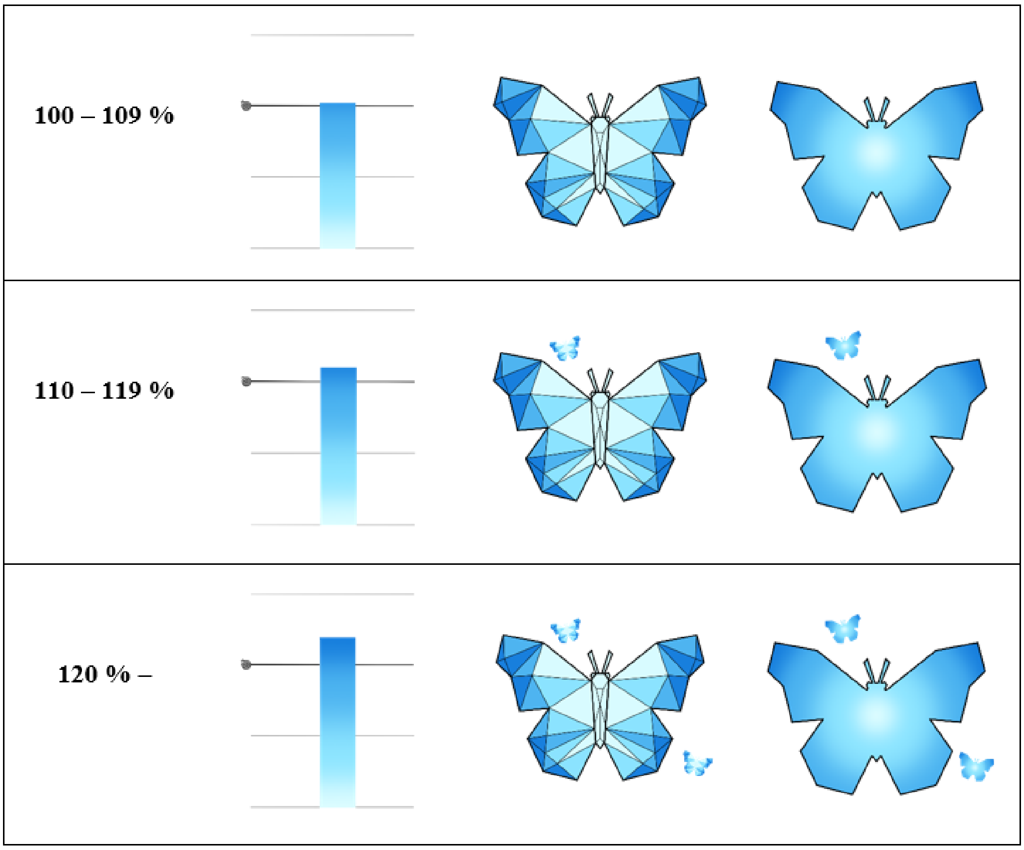
[103] Jacob O. Wobbrock and Julie A. Kientz. Research contributions in human-computer interaction. *Interactions*, 23(3):38–44, apr 2016.

A

Appendix: Visualisations

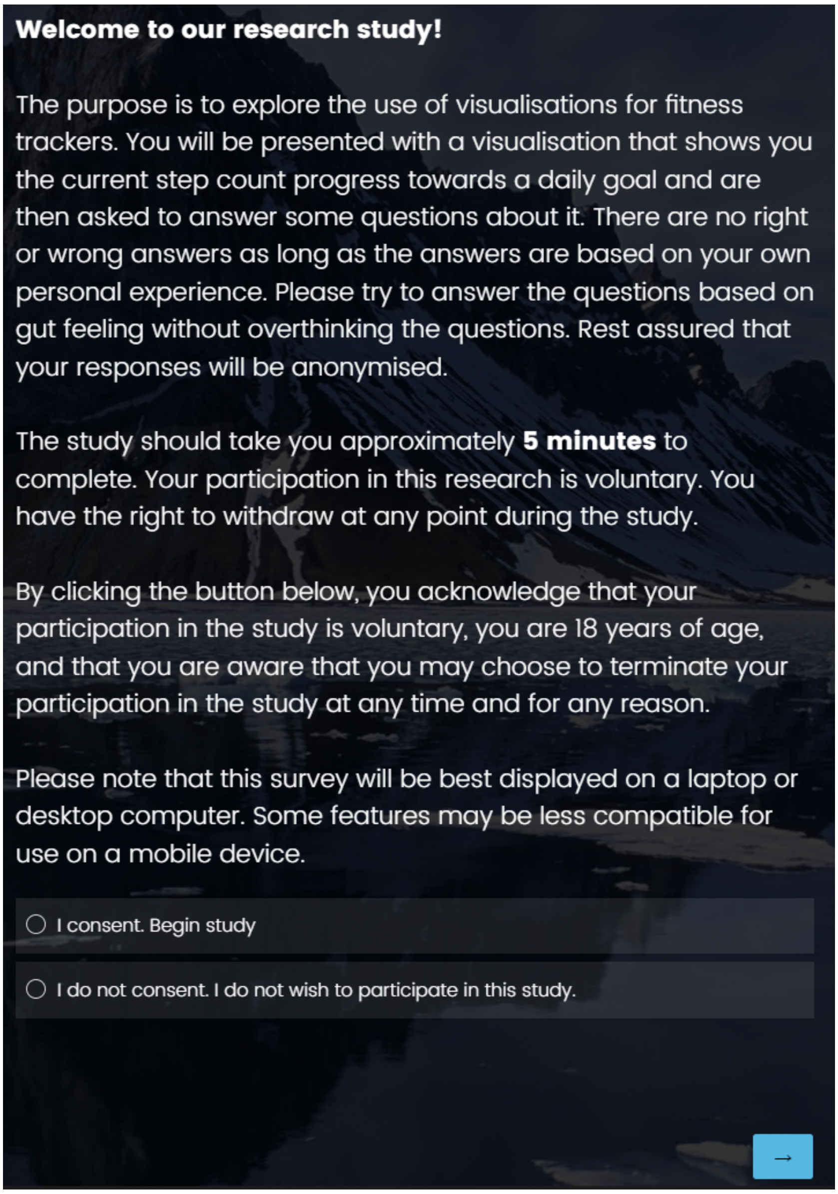


50 – 59 %	
60 – 69 %	
70 – 79 %	
80 – 89 %	
90 – 99%	


100 – 109 %
110 – 119 %
120 % –

B

Appendix: Survey



Welcome to our research study!

The purpose is to explore the use of visualisations for fitness trackers. You will be presented with a visualisation that shows you the current step count progress towards a daily goal and are then asked to answer some questions about it. There are no right or wrong answers as long as the answers are based on your own personal experience. Please try to answer the questions based on gut feeling without overthinking the questions. Rest assured that your responses will be anonymised.

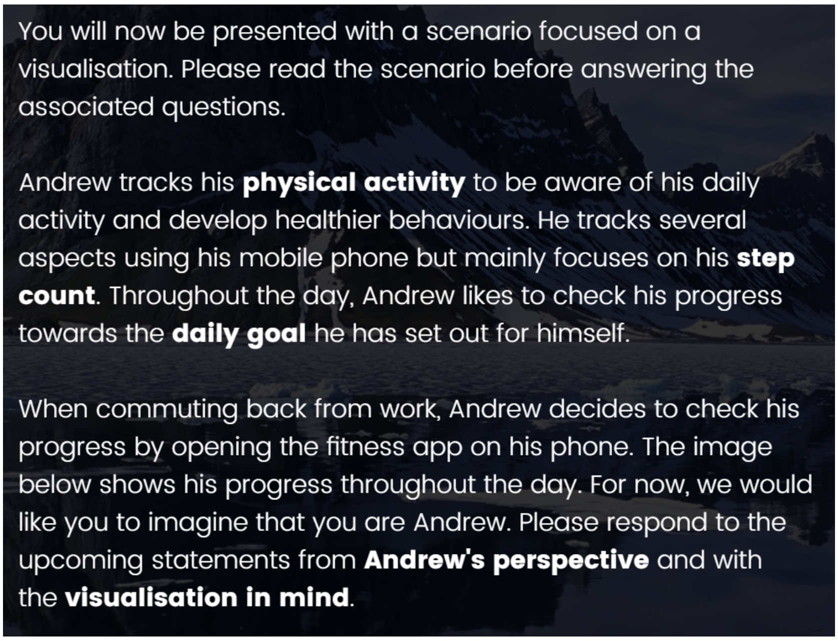
The study should take you approximately **5 minutes** to complete. Your participation in this research is voluntary. You have the right to withdraw at any point during the study.

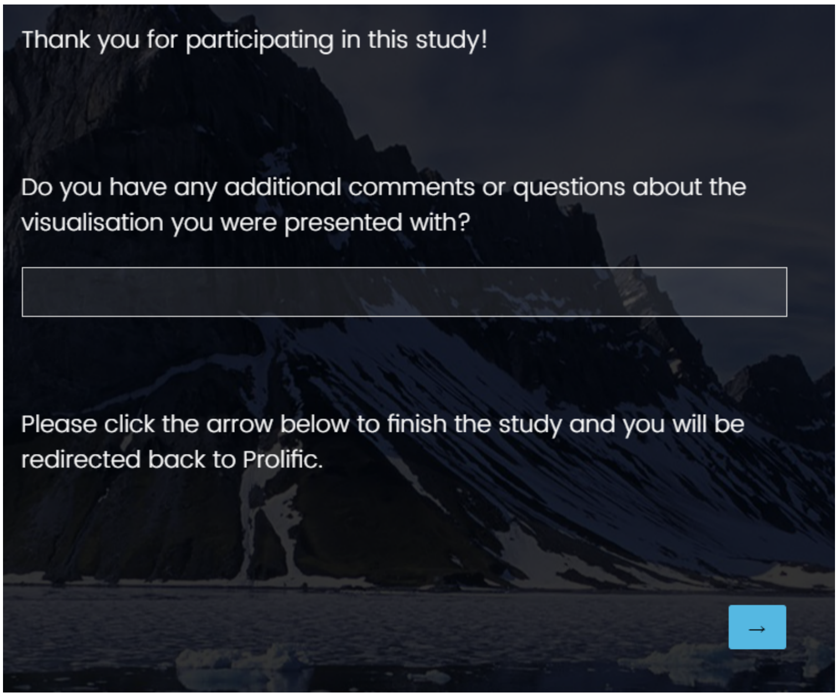
By clicking the button below, you acknowledge that your participation in the study is voluntary, you are 18 years of age, and that you are aware that you may choose to terminate your participation in the study at any time and for any reason.

Please note that this survey will be best displayed on a laptop or desktop computer. Some features may be less compatible for use on a mobile device.

- ○ I consent. Begin study
- ○ I do not consent. I do not wish to participate in this study.

→

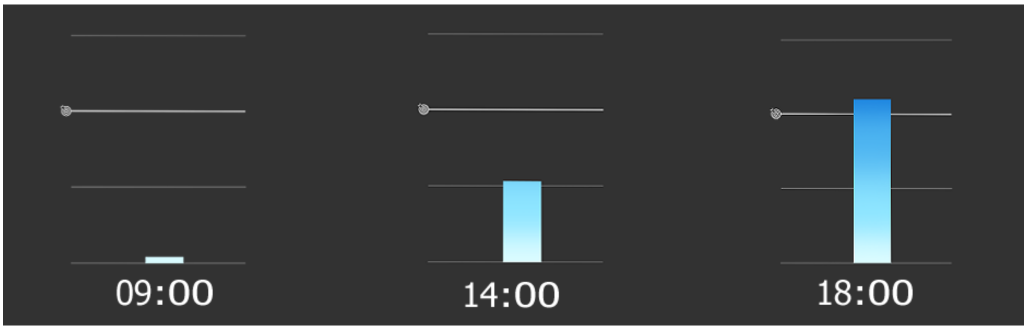

You will now be presented with a scenario focused on a visualisation. Please read the scenario before answering the associated questions.
Andrew tracks his physical activity to be aware of his daily activity and develop healthier behaviours. He tracks several aspects using his mobile phone but mainly focuses on his step count. Throughout the day, Andrew likes to check his progress towards the daily goal he has set out for himself.
When commuting back from work, Andrew decides to check his progress by opening the fitness app on his phone. The image below shows his progress throughout the day. For now, we would like you to imagine that you are Andrew. Please respond to the upcoming statements from Andrew's perspective and with the visualisation in mind.


Thank you for participating in this study!
Do you have any additional comments or questions about the visualisation you were presented with?
Please click the arrow below to finish the study and you will be redirected back to Prolific.
→

C

Appendix: Conditions

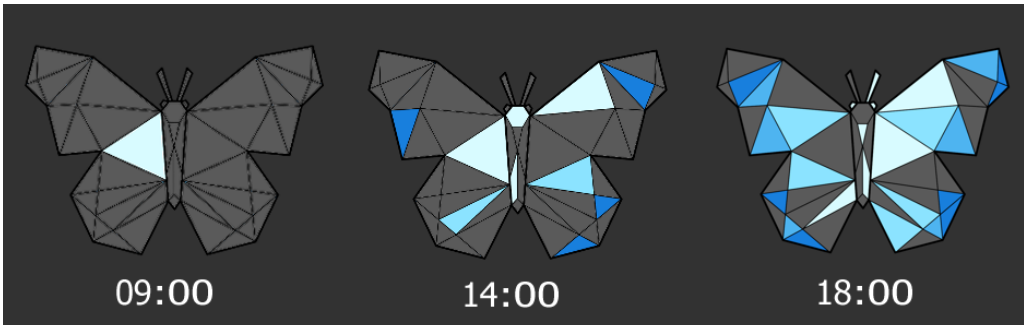
LowC



LowNC

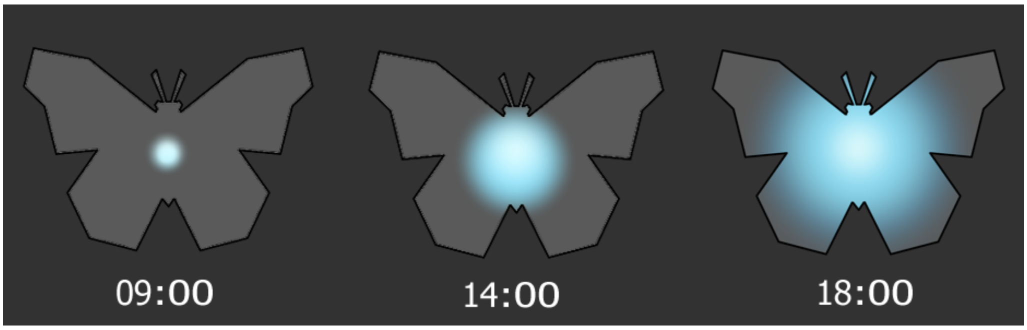
MediumC

MediumNC



HighC

HighNC



D

Appendix: Interview Protocol

Introduction

Thank you for participating in our study. This interview will be about visualisations of fitness tracking data, used in a mobile application. It will not take more than 30 minutes and will include around 15 questions regarding your experiences and feelings. Your participation in this research is voluntary and you have the right to withdraw at any point during the study. Your answers will be anonymous and if you agree to participate, the audio will be recorded for the purpose of transcribing the responses. Do you consent to participate?

Imagine that you have decided to start tracking your step count to improve your daily activity and are going to use this application to review your progress. Note that it is very simplistic, and the focus is mainly on the visualisations.

- *Participants get to try the prototype.*

- *Participants fill out the SUS questionnaire.*

Comparison

What are your initial thoughts on these visualisations?

Which visualisation do you prefer?

Why do you prefer that visualisation?

How would you feel about having access to more than one of these visualisations when using the application?

Do you see any benefits to having these multiple choices available?

Design

What do you think about the design of the bar chart?

(Show medium version) What do you think about the shape of this visualisation? How do you feel about what it represents?

What do you think about the colour of the visualisations?

Does the colour evoke any emotion in you? Why do you believe that's the case?

Interpretation

How would you interpret your current step count progress?

Do you feel like you need anything additional to better assess your progress?

How do you believe your continuous tracking would evolve, having these visualisations in mind?

(If applicable, show high version) What do you believe the smaller butterfly shapes represent?

End

Thank you for participating in this study. Do you have any additional comments or questions?