The Halo Effect on User Interfaces

An evaluation of design strategies involving aesthetics and design metaphors

Simon Lind

Simon Lind
Spring 2015
Bachelor's thesis, 15 hp
Supervisor: Lars-Erik Janlert
External Supervisor: Jonas Karppinen
Examiner: Andrew Wallace
Bachelor's programme in Computing Science, 180 hp

Abstract

This study has focused on the relation between aesthetics and usability within a user interface based on a design metaphor. Three designs with a tree metaphor and one design as a simple list has been used for evaluation. These designs have been evaluated by six usability experts with a heuristic inspection and by user testing. The heuristic inspection consisted of twelve specialized heuristics for smartphones. The user testing was done with a field-study and a think-aloud protocol. The results from both the heuristic evaluation and the user testing showed that the simple list design had the highest usability but one of the tree metaphor designs was the most aesthetically pleasing. However, the user did prefer the aesthetically pleasing design. This suggest that there is a relation between perceived aesthetics and perceived usability similar to the halo effect.

Haloeffekt inom användargränssnitt

Sammanfattning

Denna studie har fokuserat på relationen mellan utseende och användbarhet i ett användargränsnitt med en designmetafor. Tre designer med en trädmetafor och en design med en simpel lista har använts för utvärdering. Dessa designer har utvärderats av sex användbarhetsexperter med en heuristisk utvärdering samt utvärderats genom användartestning. Den heuristiska utvärderingen bestod av tolv specialiserade heuristiker för smarttelefoner. Användartestningen utfördes med en fältstudie och en tänka-högt metod. Resultatet från både den heuristiska utvärderingen och användartestningen visade att designen med en simpel lista hade högst användbarhet medan en av designerna med en trädmetafor var den som var mest estetiskt tilltalande. Användaren föredrog den estetiskt tilltalande designen vilket tyder på en relation mellan upplevelsen av estetik och upplevelsen av användbarhet liknande haloeffekten.

Acknowledgements

I wish to thank my external supervisor, Jonas Karppinen at Dohi Sweden, for the great support and for taking time to answer my questions during the whole project. I would also like to thank Anna Alnefelt and all of you at Dohi Sweden who peer reviewed and discussed my project. Thanks to Nina Ingvarsson at Umeå Kommun for answering the questions about tree inventory.

Finally, I would like to thank Lars-Erik Janlert at Umeå University for supervising, giving feedback and interesting tips of books and articles for this project.

May 2015, Umeå
Simon Lind

Contents

1 Introduction

This study is based on a tree inventory web application for smartphones developed by Dohi Sweden. Tree inventory has been done in many cities in Sweden and around the world. Tree inventory involves mapping all the trees and keep record of every individual tree over the years. This can be done within the mentioned web application.

Why is tree inventory important? Jared Diamond has covered some devastating examples which put things into perspective. Diamond emphasized that societies have collapsed because of deforestation. On the Easter Island, which is located in the Pacific ocean, all trees were chopped down by its citizens. It gradually lead to the decline of the whole civilization on that island [4]. Japan did understand the consequences of chopping down the trees on their islands. Back in 1600 they began restricting the felling of trees to ensure the survival of the forest on their island [8]. Even though the city of Umeå wouldn't collapse if all its urban trees were chopped down, the inventory web application is still a very helpful tool. It might also suit other people around the world who wants to make tree inventory easier.

Before the web application was introduced in Umeå, those who worked with tree inventory had to use a notebook and insert the data into a database later. Even before that, it involved storing the data in paper archives. The web application was a huge step in the right direction to ease the work flow but it also added functionalities which could never be accomplished with only a notebook as an inventory tool. About 150 different parameters have been specified as a standard for tree inventory, all of which should theoretically be recorded [13]. The web application made it easier to insert data directly on the location of a specific tree. At the moment, some functionalities are not yet implemented. One of the missing functionalities was to display past events for a specific tree and this will be the focus in this study. Three different designs based on a tree metaphor (Design 1, Design 2 and Design 3 in Appendix) were suggested by Dohi Sweden and a fourth design (Design 4 in Appendix) was created as a simple list to work as a baseline for the evaluation.

2 Background

Form follows function, is an expression in artifact design which means that the shape or appearance is based on how it will function [18]. For a user interface this translates to function being the usability and form as being the aesthetics. When designing graphical user interfaces instead of physical objects, the appearance and function does not come natural. It is the designer's responsibility to design and connect them. A usability researcher, Donald Norman, finds it disturbing when designers concentrate on aesthetics instead of usability. However, it has been noticed that users don't separate aesthetics and usability. Earlier studies in the field of social psychology have found that beauty associates to personal attributes. This is called the halo effect and occurs when attractive people are assumed to have a desirable personality. Tractinsky was curious if beauty could effect the perceived usability in user interfaces, similar to the halo effect. He did an experiment with an automated teller machine (ATM) and manipulated the layout of its buttons. He found a strong correlation between perceived aesthetics and perceived usability. This means that the design aesthetics did influence the perception of the usability of the system [16].

Other studies have found no correlation and also that perceived beauty is stable even after

interaction [6, 7]. Hassenzahl criticized the findings of Tractinsky by stating that the ATM's layout being too simplistic. The variation in aesthetics was only spatial and didn't include any other design dimension such as color or form. He also mentioned that hedonic stimulation can influence the evaluation. Hedonic stimulation is a term to describe how people find it interesting or stimulating with new or challenging things. Hassenzahl studied beautiful and ugly designs of a media player. The media player design which was perceived as beautiful was not perceived as more usable [6].

But it seems there is more to it. A study by Tuch et al. suggested the opposite to Tractinsky's findings but in contrast to Hassenzahl found perceived aesthetics being affected by interaction. The study involved a web shop where they manipulated the background as the aesthetic parameter and the navigation as the usability parameter. Tuch et al. suggested that the perceived usability and aesthetics could be reversed based on the user's experience. This would mean that the perceived aesthetics would get higher after interaction, if the usability is high [17].

Hassenzahl and Monk have reviewed 15 studies on perceived beauty with perceived usability. They summarized the findings by pointing out the wide variety of design parameters used to study the beauty and usability relation. They suggest paying more attention to which method and analysis is used to find the underlying process [7].

Another aspect of design is using metaphors as a design tool. When designing a new system, using a metaphor can help users understand and remember how things work. As computers became a consumer product rather than just a professional tool, making it easier to understand and learn became important. A metaphor has the potential to make a complex product more understandable, learnable and appealing [11].

Metaphors are fundamental for the human language and how we act. Metaphors are much more than just words as they structure concepts and they are often used to transfer a concrete experience to something abstract. For example, when time is used as a money metaphor, you can use the phrases ”spend time”, ”waste time” or ”run out of time” and therefore people can picture it. Metaphors are also used with an orientation where up means glad and down means sad and the future is ahead and the past is behind. Metaphors have a great impact on how we view the world [10]. Unfortunately, there are some metaphors which have cultural differences. A simple example when the metaphor differs in some areas in the world is when an orientation is based on future and past events. Different cultures will perceive future to the left or to the right. Western cultures read from left to right whereas text written in Hebrew is read from right to left [12]. This needs to be taken into account when using an orientation metaphor for presenting past events.

A metaphor is not always the answer to create a great system. It comes with the risk of creating usability problems and constraints due to the metaphor [3, 9]. Metaphors are therefore a double edged sword. Hudson mentions that metaphors are only helpful when learning a new system. A classic example is the desktop metaphor in computers where office workers could quickly adapt to computers. But there are lots of examples where the metaphor also lead to confusion. In an early Macintosh version, if you wanted to eject a floppy disk you would have to drag it to the trash can and that didn't match the user's mental model [9]. Another example was the typewriter metaphor as a word processor. The user made many errors because of the mismatch between a typewriter and a word processor and the errors were simply the result of the users' expectations. This is not a problem any more because most users today have likely never used a typewriter, but the mismatched expectation is still

the biggest issue when using a metaphor in a design [2]. It is not surprising finding errors in a design metaphor, because metaphors by definition are inexact. It becomes a problem when users take them too literal [2, 9]. Anyhow, Norman explains that when changing a metaphor people will at the beginning get confused but essentially adapt to the change [12]. Using a system for a while will make the user understand which parts of the metaphor work and which don't.

Metaphors is the key to learn new concepts [2]. For a user interface, metaphors are an important design tool but must be carefully used [9]. Choosing the right metaphor can be difficult but rewarding when done correctly.

2.1 Purpose

The aim of this study is to find out if the aesthetics can affect the perceived usability. In order to do this, the four designs first have to be evaluated to determine which design has the highest usability and which is the most aesthetically pleasing. Then, let users interact with the designs and examine whether a user prefers a design with the highest usability or a design which is more aesthetically pleasing.

Three out of four designs have a tree metaphor for displaying past events. It will therefore be valuable to evaluate whether the tree metaphor is useful in understanding the design or if the metaphor creates a mismatch between the user's mental model and the interaction.

The findings in this study should be helpful when designing a system with a graphical user interface.

3 Method

3.1 Theory Behind The Evaluation Methods

Heuristic inspection is a commonly used evaluation method among interaction designers. It is used to detect usability problems with a user interface. About five usability experts is recommended. They systematically inspect the user interface as they rate and comment each heuristic. All heuristics are rated with 1 to 10, where 1 is the worst and the problem must be fixed and 10 is the best and works fine. Jakob Nielsen have recommended ten heuristics which are often used in a heuristic inspection. A heuristic inspection is an easier and cheaper method of evaluation because it doesn't involve finding users. Worth to note is that it doesn't replace user testing but is rather a complement to find obvious design flaws and usability problems [14, 15]. A limitation with the standard heuristics is that they don't take smartphones into account. Instead of strictly using the standard heuristics recommended by Nielsen, a better approach is to make a list of heuristics based on usability guidelines for smartphones [1]. Therefore, a modification of the standard heuristics was needed in order to suit this evaluation.

User testing is the most important aspect when evaluating a design. A real user might identify problems while performing a realistic task that a designer had never thought of. Think-aloud protocols is a useful method during user testing in combination with user observation. It involves the user constantly expressing every thought that comes to mind while examining an interface. The advantage of this method is that it can reveal problems or sur-

prises as the user explores the interface. For some users, when they constantly talk, it might instead lead to a distraction or they feel it being unnatural [15].

3.2 Heuristic Inspection

Participants. Six usability experts (three men and three women) which have knowledge and experience using the heuristic inspection method. All experts have at least a bachelor degree in cognitive science. All experts were recruited by email and asked if they would like to participate voluntarily. The recommendation of five participants makes six participating usability experts sufficient for the heuristic inspection.

Equipment. Four interactive mock-ups, based on the designs, were assembled for the heuristic inspection (see Appendix A). The mock-ups consisted of 44 static images made interactive by defining clickable areas which would change the current image. The mock-ups were available on both a smartphone and a smartphone emulator on a computer. The experts used their own smartphones and computers. One expert used the smartphone emulator. A form of heuristics was created for each mock-up and the complete list of heuristics used for the heuristic inspection can be seen in Table 1.

Procedure. The inspection took between 45 to 120 minutes to complete. The variance was due to the absence of a time limit and the spare time available for the experts. The experts began by reading the instructions. They were then advised to read all the heuristics before examining the mock-ups. The experts could choose the order of which mock-up to inspect. They systematically inspected the mock-ups as they rated and commented each heuristic. The heuristics form was answered in parallel with the testing of the mock-ups and the mock-ups were evaluated with a smartphone or a smartphone emulator. After the heuristic inspection was completed the experts were asked to choose which design they preferred.

Table 1: The heuristics with description and a space for rating 1 to 10 and comments

Heuristic	Description	Rating	Comments
Visibility of system status			
Current status	Is it easy to know where you are and where you are going?		
Match between system and the real world			
Language	Appropriate use of words and phrases. No technical terms.		
Mental model	Does it work as expected? Does the navigation and buttons remind of similar applications?		
Metaphor	How does the tree metaphor work in the application?		

Continued on next page

Table 1 – continued from previous page

Heuristic	Description	Rating	Comments
Consistency and standards			
Consequent functionality	Is the behavior for buttons, functions and appearance consequent?		
Platform	Does the design and button placement follow the convention for the targeted platform?		
Recognition rather than recall			
Minimal cognitive load	The user should not have to remember how to execute a function. The user should recognize similar actions.		
Flexibility and efficiency of use			
Shortcuts	The expert user should get to the desired location/information efficiently and effectively.		
Readability	Is the line spacing, contrast, images and colors optimal for reading?		
Aesthetic and minimalist design			
Aesthetically pleasing	Does the design enhance the user experience and interaction?		
Relevant information	Only necessary text and graphical elements is presented.		
Error Prevention			
Physical interaction and ergonomics	The application is supposed to be used with a smartphone where the display is relatively small. Therefore, it is important to prevent the user from selecting the wrong item if buttons are small or close together. This is also known as the fat-finger error.		

3.3 User Testing

Participants. One user with over seven years of tree inventory experience and prior experience of using the tree inventory web application by Dohi Sweden. Only a few people have tree inventory as a profession in Sweden and not everyone of them have used the tree inventory web application. Thus, the participating user was the most suitable for user testing.

Equipment. Two prototypes were developed for the user testing (see Appendix B). One prototype was based on the design with the highest rating of aesthetically pleasing in the heuristic inspection. The second prototype was based on the simple list. The prototypes were tested on an iPhone 6 which has a 4.7 inch screen and a 750x1334 pixel resolution. Two forms were created for the user testing. The first form contained questions of the user's background and general questions about tree inventory. The second form contained questions about the actual test of the prototypes. The aim of the second form was to address which design was the most aesthetically pleasing and which design the user actually preferred. The second form also gave the opportunity to write additional comments or suggestions for improvements.

Environment. The location for the user observation was chosen to represent a work situation in an environment which was familiar to the user. The location was outside close to one of the trees beside a street and nearby houses. The answering of the forms and the discussion were held in an office.

Pilot. The prototypes were peer-reviewed by colleagues at Dohi Sweden to find problems within the prototypes. This was made in order to ensure that the prototypes didn't have any issues which could affect the user testing.

Procedure. The user answered the first form which contained questions about the user's background. The user was then moved to the testing environment outside. The user was given the instructions to think-aloud and try to express every detail as much as possible. The user was given specific tasks, but the user also had to come up with a realistic example to perform. The tasks were repeated for both prototypes. The second form was answered after the user observation. Both forms were a type of semi-structured interview where each question was encouraged to be discussed. A few questions was asked in post-discussion to let the user express any thoughts not covered in the forms. Notes were taken during the whole session.

4 Result

4.1 Heuristic Inspection

The average of the six experts' ratings was calculated for each heuristic. The average on each design was calculated for the overall ratings. The averages of the ratings from the heuristic inspection can be seen in Table 2. In addition to the ratings, the experts proposed a lot of improvements to each design. Design 4 got the highest overall rating average and the highest average rating in all individual heuristic except the Language and Aesthetically pleasing heuristics. Design 3 got the highest average rating of the language heuristic. Design 1 got the highest average rating of the aesthetically pleasing heuristic.

Table 2 The result of the rating 1 to 10 from the Heuristic Inspection

	Design 1	Design 2	Design 3	Design 4
Visibility of system status				
Current status	6,3	6,8	5,2	**8,3**
Match between system and the real world				
Language	8,3	9,0	**9,2**	8,2
Mental model	6,8	6,3	5,7	**9,2**
Metaphor	5,5	6,8	4,3	**8,2**
Consistency and standards				
Consequent functionality	8,7	7,8	7,5	**9,2**
Platform	7,7	5,8	8,0	**8,8**
Recognition rather than recall				
Minimal cognitive load	8,5	8,0	7,8	**8,8**
Flexibility and efficiency of use				
Shortcuts	6,8	6,7	6,8	**8,2**
Readability	8,3	7,7	6,7	**9,3**
Aesthetic and minimalist design				
Aesthetically pleasing	**8,7**	7,8	5,7	7,0
Relevant information	8,5	7,3	6,8	**9,2**
Error Prevention				
Fysical interaction och ergonomics	8,5	7,0	8,2	**9,2**
	7,7	**7,3**	**6,8**	**8,6**

When the experts were asked to choose which design they preferred, two experts explained that they liked Design 1 because of its aesthetically pleasing design. Two experts liked the idea of Design 2 but pointed out that it needs modifications. Three experts said that the metaphor designs were more fun to use. But five out of six experts clearly stated that Design

4 probably is easier to use and best suited for the application.

4.2 User Testing

Four usability problems were observed during the user observation of Design 1 and the data interpretation can be found in Table 3. Each usability defect was rated as low, medium or high, depending on the severity of the observed error. All the usability defects where the user selected something else than the target, were rated as high. The usability defects which were rated as medium, did not cause the user to make an error but might become a potential risk of missing information or take unnecessary long time to perform. The user didn't ask how to operate any of the two prototypes and was quick to initiate the first task. The user expressed that it was easy to understand the icons and how to navigate.

Table 3 Data interpretation form for the user observation of Design 1

Usability Observation	Evaluator's comment	Cause of the usability defect	Severity rating
The user selected the wrong icon	The user selected the first occurring icon which were similar to the target	The icon was semantically similar. The design doesn't have a label visible along with the icon.	High
The user selected the wrong month	The user was quite confident until opening one event and found it was the wrong month	The time line displaying months didn't match the mental model of the user.	High
The user hesitated whether to scroll up or down after using the filtering function	The user looked surprised when all the icons disappeared after using the filtering function.	The filter removes all unnecessary information and leaves the selected ones. In this case the targeted icon wasn't in the current screen view.	Medium
The user must scroll to see the information if the icon is close to the bottom	It is intuitive to scroll down and tap the icon close to the bottom for easier reach	The information is displayed below the icon after it has been tapped. In this case the information was just below the current screen view.	Medium

There were no noticeable usability defects during the observation of design 4. The user made a comment that "It looks compact" and expressed that it would be better if the events were directly proportional to time as they occurred.

In the second form the user stated that Design 1 was the most aesthetically pleasing. The user preferred Design 1 but made comments of improvements.

The user gave design suggestions for both Design 1 and Design 4. One of the things was

the ability to search for a specific parameter. Other suggestions were outside the scope of this study.

5 Discussion

The heuristic inspection showed that Design 1 was the most aesthetically pleasing while Design 4 got the highest overall usability rating. The same pattern was found in the user testing where the user thought Design 1 was the most aesthetically pleasing. But the difference between the heuristic inspection was that the user preferred Design 1 even though Design 4 had no observed usability defects. This result supports Tractinsky's findings that aesthetics affects the perceived usability [16]. It is also possible that Design 1 gained the halo effect by being aesthetically pleasing.

The user might have preferred Design 1 based on its appearance and that phenomenon has been noticed in the music industry as well. For example, sound engineers nowadays use a lot of plug-ins and virtual sound processors. When the graphical user interfaces are an imitation of physical recording equipments, the sound engineers tend to "listen with their eyes" instead of their ears when buying a new product. However, a good looking interface can make the user feel enjoyed when working with the product [5].

Three usability experts said that the tree metaphor was fun to use instead of a simple list. The user testing showed that the tree metaphor helped the user quickly understand the design. The metaphor seemed to have enhanced the experience and gave the notion that it was fun to use. This could also have had an impact on why the user preferred the metaphor design instead of the simple list design. The tree metaphor and the actual task of working with past events for trees has by itself a fun connection. To spin this even further, it may also be a matter of hedonic stimulation where the metaphor design can be seen as new or challenging [6].

The application was developed for professional use which means that it might be used a lot. Suppose that the user worked with Design 1, which had some usability defects, for longer use, the usability might become a bigger problem and maybe it won't be perceived as aesthetically pleasing any more. The user testing might only have captured the first impression.

6 Conclusion

This study supports the theory that aesthetics affects the perceived usability. Even if the design had more usability defects the user still preferred the design which was more aesthetically pleasing. This could mean that a user interface can obtain a halo effect and that the aesthetics has a great impact on our first impression of a system. In addition to this finding, a design metaphor can help the user to understand and learn a new system and also make the system more fun to interact with.

How can this finding be useful when designing a system with a graphical user interface? We should aim to have it all. Extraordinary usability and amazing aesthetics built upon a well thought-out metaphor and we will succeed without a doubt. However, if we are only

going to present a prototype for the first time to the users, we should make sure to make the design aesthetically pleasing which might obtain a halo effect. The design should also be based upon a metaphor which makes the interaction more fun. If the users then prefer the design, we could continue with the focus on the usability.

6.1 Further Research

Limitations. The usability experts didn't have any domain knowledge in tree inventory prior to the heuristic inspection and that may have altered the usability ratings. There was only one available user. If there had been additional users it would have been possible to analyze statistical effects, such as age, completion rates or time-on-task which can be a factor.

Another limitation in this study was that the observer also took notes. That contributes to the risk of missing anything important during the observation, because the focus shifts between the notebook and the observed user. This could have been solved by video filming the user testing in addition to taking notes.

Future suggestion. The user testing might have captured the first impression. Design 1 and Design 4 could be evaluated before and after, using it for a longer period of time or at least multiple times with the same user. If the halo effect is the answer to why users choose a design with lower usability, one could investigate how bad the usability could get before the users start abandoning the system. Time, repetition and usability could be manipulated to test this.

A tree metaphor was used for all but Design 4. It would be interesting to add another design with a different metaphor to examine the effect the metaphor had to the perceived aesthetics and perceived usability.

As shown in this study, the aesthetics affected the perceived usability. Future studies should try to find the core of the aesthetics–usability relation and which parameters are important.

References

[1] 4ourth Mobile. Heuristics for evaluating mobile designs. http://4ourth.com/wiki/Heuristics%20for%20Evaluating%20Mobile%20Designs, 2014. [Online; accessed 21-May-2015].

[2] Pippin Barr. User-interface metaphors in theory and practice. 2003.

[3] Ward Mitchell Cates. Systematic selection and implementation of graphical user interface metaphors. *Computers & Education*, 38(4):385–397, 2002.

[4] Jared Diamond. *Collapse: How societies choose to fail or succeed.* Penguin, 2005.

[5] Nils Eriksson. Lyssnar du med ögonen? http://bloggar.m3.idg.se/musikproduktionsbloggen/lyssnar-du-med-ogonen/, 2015. [Online; accessed 24-May-2015].

[6] Marc Hassenzahl. The interplay of beauty, goodness, and usability in interactive products. *Human-Computer Interaction*, 19(4):319–349, 2004.

[7] Marc Hassenzahl and Andrew Monk. The inference of perceived usability from beauty. *Human–Computer Interaction*, 25(3):235–260, 2010.

[8] Peter Holmgren, Kailash Govil, and Dan Altrell. Brief on national forest inventory nfi, japan. *Food and Agriculture Organization of the United Nations*, 19:1–16, 2007.

[9] William Hudson. The whiteboard: metaphor: a double-edged sword. *interactions*, 7(3):11–15, 2000.

[10] George Lakoff and Mark Johnson. *Metaphors we live by*. University of Chicago press, 2008.

[11] Aaron Marcus. Human communications issues in advanced uis. *Communications of the ACM*, 36(4):100–109, 1993.

[12] Donald A Norman. *The Design of Everyday Things: Revised and Expanded Edition*. MIT Press, 2013.

[13] Johan Östberg, Tim Delshammar, Ann-Mari Fransson, and Anders Busse Nielsen. Standardiserad trädinventering och insamling av data för träd i urban miljö. 2012.

[14] Yvonne Rogers, Helen Sharp, and Jenny Preece. *Interaction design: beyond human-computer interaction*. John Wiley & Sons, 2011.

[15] Debbie Stone, Caroline Jarrett, Mark Woodroffe, and Shailey Minocha. *User interface design and evaluation*. Morgan Kaufmann, 2005.

[16] Noam Tractinsky, Adi S Katz, and Dror Ikar. What is beautiful is usable. *Interacting with computers*, 13(2):127–145, 2000.

[17] Alexandre N Tuch, Sandra P Roth, Kasper Hornbæk, Klaus Opwis, and Javier A Bargas-Avila. Is beautiful really usable? toward understanding the relation between usability, aesthetics, and affect in hci. *Computers in Human Behavior*, 28(5):1596–1607, 2012.

[18] Wikipedia. Form follows function. http://en.wikipedia.org/wiki/Form_follows_function. [Online; accessed 22-May-2015].

A Appendix: Mock-ups

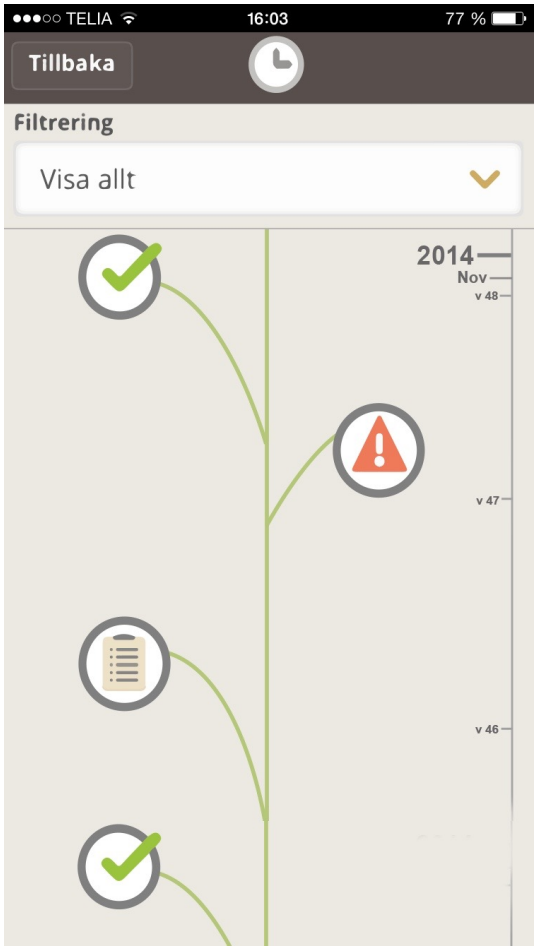

Figure 1: Mock-up of Design 1

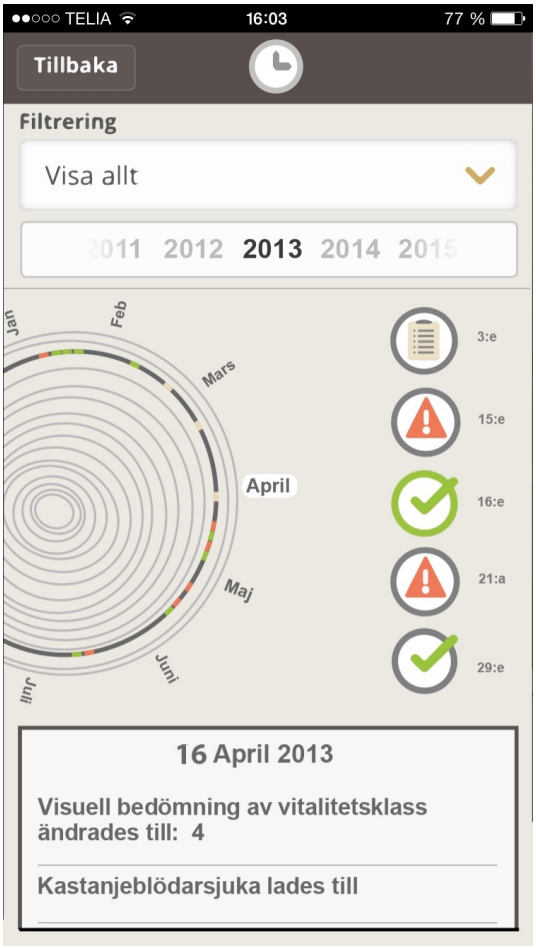

Figure 2: Mock-up of Design 2

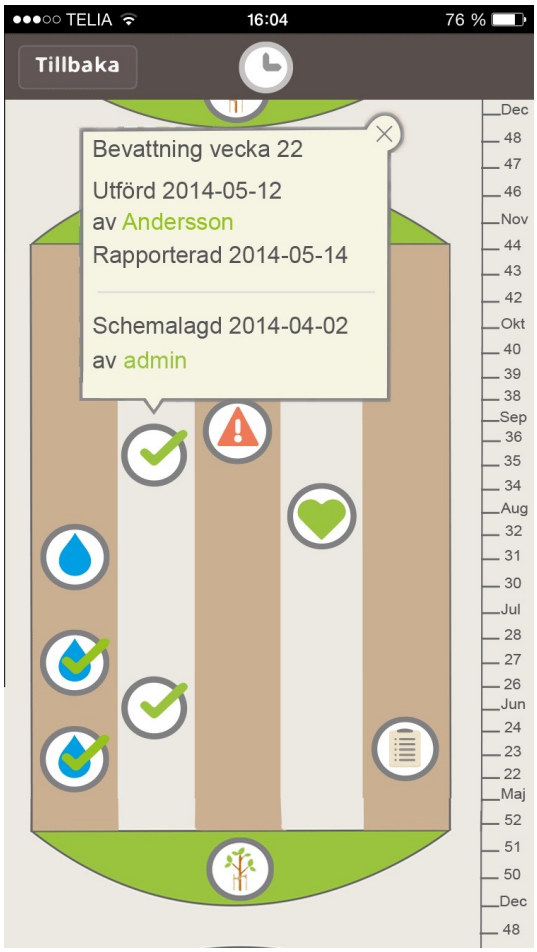

Figure 3: Mock-up of Design 3

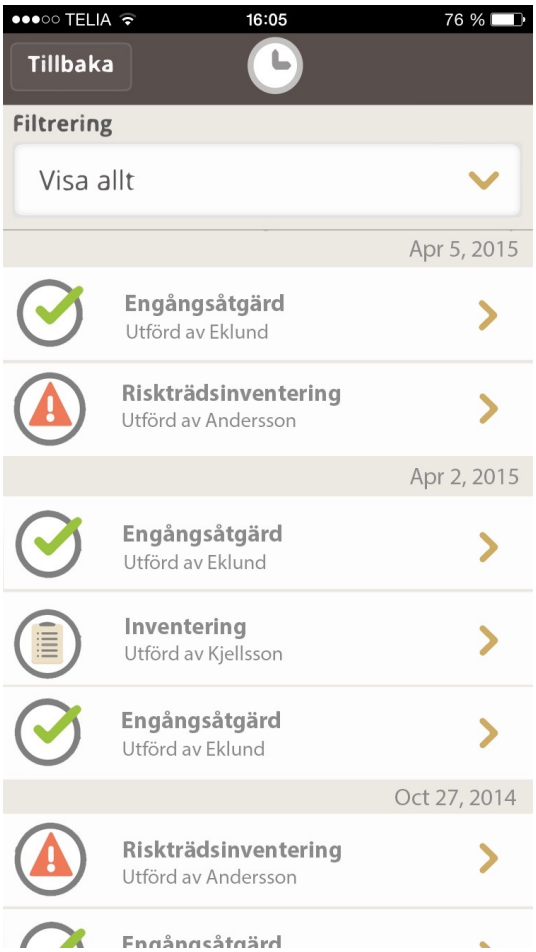

Figure 4: Mock-up of Design 4

B Appendix: Prototypes

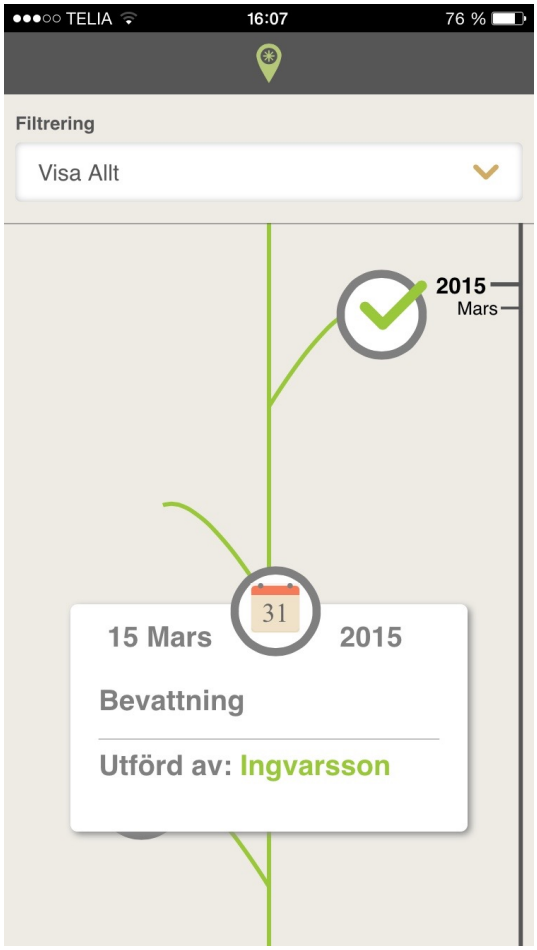

Figure 5: Prototype of Design 1

Figure 6: Prototype of Design 4