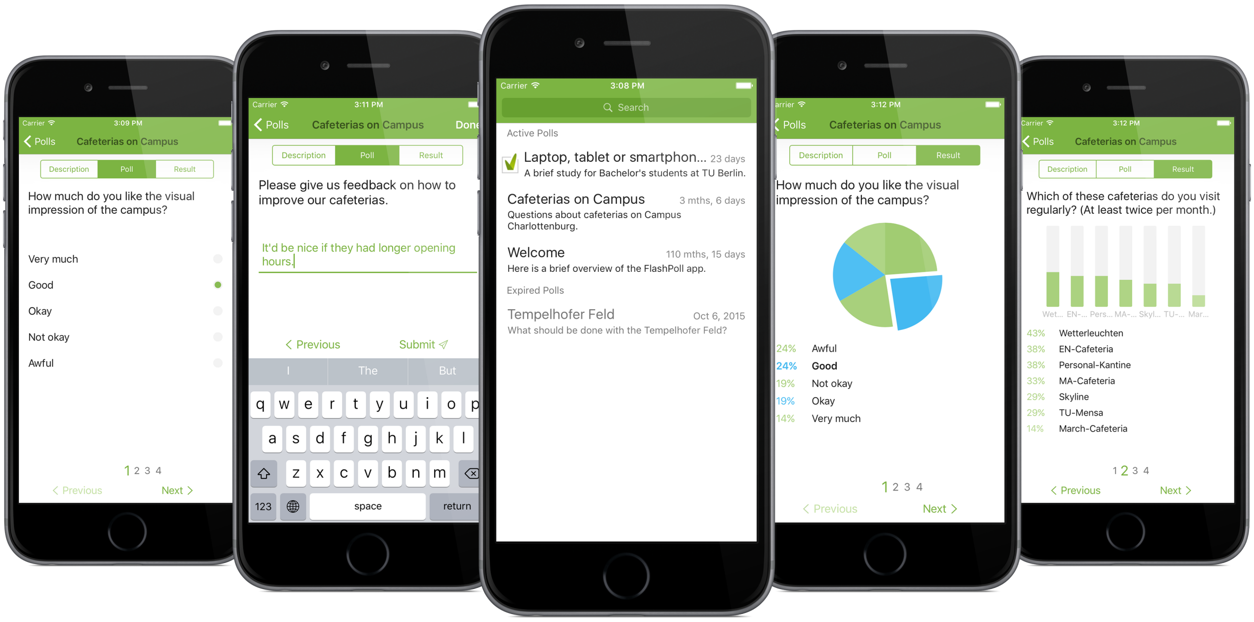
CHALMERS
UNIVERSITY OF TECHNOLOGY

UNIVERSITY OF GOTHENBURG



FlashPoll: A Location-aware Polling Application for iPhone

Examining Differences Between iOS and Android Users in the Context of Mobile Polling Applications

Master's Thesis in Interaction Design & Technologies

ERICH GRUNEWALD

Department of Applied Information Technology
CHALMERS UNIVERSITY OF TECHNOLOGY
UNIVERSITY OF GOTHENBURG

Gothenburg, Sweden 2017

MASTER'S THESIS 2017:003

FlashPoll: A Location-aware Polling Application for iPhone

Examining Differences Between iOS and Android Users in the Context of Mobile Polling Applications

ERICH GRUNEWALD

Department of Applied Information Technology
Division of Interaction Design
CHALMERS UNIVERSITY OF TECHNOLOGY
UNIVERSITY OF GOTHENBURG
Gothenburg, Sweden 2017

FlashPoll: A Location-aware Polling Application for iPhone
Examining Differences Between iOS and Android Users in the Context of Mobile Polling Applications
ERICH GRUNEWALD

Supervisor: Maija Poikela, Quality and Usability Lab, Telekom Innovation Laboratories, TU Berlin
Supervisor: Olof Torgersson, Department of Applied IT, Chalmers University of Technology
Examiner: Staffan Björk, Department of Applied IT, Chalmers University of Technology

Master's Thesis 2017:003
Department of Applied Information Technology
Division of Interaction Design
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: The FlashPoll app for iOS.

Typeset in LaTeX
Gothenburg, Sweden 2017

FlashPoll: A Location-aware Polling Application for iPhone
Examining Differences Between iOS and Android Users in the Context of Mobile Polling Applications
ERICH GRUNEWALD
Department of Applied Information Technology
Chalmers University of Technology

Abstract

There exists a wealth of methods to make possible the learning of people's opinions, values and desires – including polls, surveys, studies, focus groups and so on. The FlashPoll project, of which this thesis was a part, aims to shorten the distance between citizens and those who wish to learn about them. It works by letting such a person – an administrator or decision-maker – create a poll and assign it to a geographical area, wherein citizens who use the FlashPoll app will be able to see it and respond to it.

This thesis was concerned both with how best to design the interaction of a location-based mobile polling application for iOS; and with how iPhone and Android users differ in terms of attitude, usage patterns and so on, in the context of such an app.

I designed and developed an iPhone version of the FlashPoll tool (an Android version having already been made). Using these two versions of the FlashPoll app, I conducted a user study, wherein 40 participants (22 Android users and 18 iPhone users) spent a week with the app, reporting their experiences through daily diary entries.

The results of the thesis consisted of the completed FlashPoll iOS app itself, of the various findings from the user study, and – drawing upon the other two – of a set of eight guidelines for future designs related to mobile polling.

Keywords: interaction design, design guidelines, public opinion, opinion poll, mobile polling, electronic survey, smartphones, iPhone, iOS, Android.

Acknowledgements

This Master's thesis was a part of the Interaction Design & Technologies programme at Chalmers University of Technology. The work itself was carried out at the Quality & Usability Lab, a part of the Telekom Innovation Laboratories, in Berlin, Germany.

I would like to thank T-Labs in general for having welcomed me, and Maija Poikela in particular for having supervised my work and provided a wealth of knowledge and experience for me to draw upon. What is good in this thesis is largely thanks to Maija, and what isn't, can safely be said to originate with the author.

I am indebted to Robert Schmidt, whose knowledge was invaluable throughout, and who helped me with translations and with conducting the user study; and Felix Kaiser, who also assisted with translations and with the study.

Finally, I want thank Olof Torgersson at Chalmers for his help in providing a direction for the thesis, and for his feedback on this report.

Erich Grunewald, Gothenburg, November 2015

Contents

List of Figures

List of Tables

1

Introduction

This thesis concerns the creation of an iOS app for location-based polling of public opinion. It aims to investigate, in this context, differences in attitudes and behaviours between Android and iOS users, and moreover what the implications of these differences are for future designs.

1.1 Public opinion

“No feats of heroism are needed to achieve the greatest and most important changes in the existence of humanity; neither the armament of millions of soldiers, nor the construction of new roads and machines, nor the arrangement of exhibitions, nor the organization of workmen’s unions, nor revolutions, nor barricades, nor explosions, nor the perfection of aerial navigation; but a change in public opinion.”

So wrote Count Lev Nikolayevich Tolstoy in his essay *Christianity and Patriotism* as the 19th century drew to a close (Tolstoy 1967). The world is now as it was then filled with people who have opinions, values and desires – the human range of them is limited only by our capacity to describe them. But how does one get a sense of those opinions, values and desires? And how does one know which of those opinions, values and desires are relevant to the task at hand?

The flow of information in society, which goes both upward and downward, has, since the invention of writing, been deeply influenced by technology. Today, there is a wealth of methods available to facilitate the upward flow of information, as people in positions of power use polls, surveys and studies to inform their decisions, and as corporations and political campaigns conduct focus groups in order to best design their products and direct their marketing.

The demand for these services is not likely to go down. “Public opinion is, in the first place, not something that is constant, unchanging, and stable. On the contrary, it is something that is eternally changing, moving together with the motion of humanity” (Tolstoy 1967). This inconstancy, changeability and instability makes necessary the continual measuring of public opinion, and with that necessity comes the need for easier, more robust and more accessible methods of measurement.

1.2 The FlashPoll project

The *FlashPoll* project aims make the flow of information easier, to shorten the distance between decision-maker and citizen by developing “a tool that allows for better, more satisfying and more consistent communication between citizens and administrators” (FlashPoll 2014). It works as following: an administrator (or decision-maker) creates a poll[1] through a web interface and delineates a geographical area wherein people (citizens) will be able to see the poll. The people in that area who have installed one of the FlashPoll mobile apps are notified of a new poll’s having appeared, and can then respond to that poll. Both the respondents and the administrator can evaluate the results of the poll in the form of charts and numbers. After a set period of time, the poll is closed. The administrator can then take the opinions provided by the poll into account when carrying out his or her task or when coming to some sort of a decision.

An example will serve: a librarian, wanting to find out about the library’s visitors and what they think about its service, resources and so on, uses the FlashPoll web interface to design a poll with a number of questions, perhaps related to how often one visits the library, what sort of books or magazines one is interested in, etc. The librarian then assigns the poll to a geographical area – the library – and sets it to start and end on certain dates. While the poll is active, people who use the library and have the FlashPoll app get a notification as they enter the library building, and can respond to the poll on their devices. Both the librarian and the visitors can then look at the results of the poll.

The FlashPoll project is run by a constellation of partners, including research labs, universities and companies (FlashPoll 2014). So far, it has been put to use in several real-world contexts, such as at a suburban school in Sweden, during a public debate in France and at the *Long Night of Science* event in Berlin (FlashPoll Project Consortium 2015). The project is run and funded within the framework of a so-called “innovation action line” at the *European Institute of Innovation and Technology’s* ICT Labs. This innovation action line, or programme, is called *Urban Life and Mobility*, and runs from January 2013 to the close of 2015 (FlashPoll 2014).

It may at this point be useful to make the distinction between a *poll* and a *survey*. A poll is defined here as (1) an assessment of public opinion or experience by questioning a portion of the relevant population. A survey, on the other hand, is a specific method: it is (2) a structured questioning of the public opinion/experience of a sample. However, in the context of the FlashPoll project, the term “poll” is sometimes used to cover both definitions.

The FlashPoll tools, it should be noted, are not intended for voting but rather for gathering information, and by so doing to improve communication between citizens and administrators.[2] It aims to create a direct link between “respondent, questioner

[1]This may, perhaps, be more accurately termed a survey (cf. definitions (1) and (2) later in Section 1.2), but because the FlashPoll project uses the term “poll”, this paper, too, will use that term when discussing the FlashPoll apps.

[2]It is different in this way from alternatives such as the Google Opinion Rewards app, which

and the subject of the questions", to make the answering of polls faster, and to give participating users better feedback (FlashPoll 2014). In order to reach this aim, the FlashPoll project has developed a set of tools with which a series of test runs, some mentioned above, have been carried out in the cities of Berlin, Paris, Nantes and Stockholm (FlashPoll 2014).

The FlashPoll app for Android was made in 2013 and used in a user study that same year. As was noted in the beginning of this chapter, this thesis concerns the creation of an iOS version of the FlashPoll tool, as well as a subsequent user study, which uses the two apps to investigate differences in characteristics and behaviour among Android and iPhone users.

1.3 Android and iOS

Since the introduction in 2007 of the first-generation iPhone and the unveiling of Android in the same year, the market for mobile[3] devices has been dominated by those two platforms: Google's Android and Apple's iOS. These two, although they have been around for about as long and share many of the same goals, have certain differences when it comes to how they present information to the user and how they recommend an app should be designed.

The magnitudes of these differences have varied since the platforms' entrances on the market. In 2012, Apple announced that Scott Forstall, its senior vice president, would be leaving the company (Apple 2012). Forstall, who led the software development team for the first iPhone, had along with Steve Jobs been a proponent of skeuomorphic[4] design at Apple (Wingfield and Bilton 2012). With his departure, the use of skeuomorphism was reduced, and at the Apple Worldwide Developers Conference in 2013 Jonathan Ive, who had then recently taken on the role of leading Apple's interface design, introduced the flat and simple design of the new iOS 7 (Kuang 2013). The new design guidelines place emphasis on clear, spacious design where interface defers to content and on the use of colour rather than texture to indicate interactivity (Apple 2015b).

In a way, Android has been moving in the opposite direction. At the Google I/O

was introduced by Google in 2013 and allows its users to answer surveys (constructed and paid for by businesses) in exchange for Google Play credits (Google 2015b). Although there are some similarities, the FlashPoll project, unlike Google Opinion Rwards, is made as part of a research project, with public administrators, not private businesses, in mind.

[3] Throughout this thesis, the nouns "mobile" and "mobile device" are used interchangeably, and generally refer to smartphones specifically.

[4] Skeuomorphism in interaction design refers to graphical, auditory or interactive elements which mimic or resemble physical counterparts – a sort of metaphor. Cooper et al. (2007) refers to this as "mechanical-age representations". Dunne (2008) comments that, in the mimetic approach, "[nothing] is what it appears, but simply an allusion to something we are already familiar with". Neal Stephenson, in his essay *In the Beginning ... Was the Command Line*, compares early graphical user interfaces, with their windows, desktops, documents and other such metaphors, to Disney, which, so he writes, "does most of its communication without resorting to words, and for the most part, the words aren't missed" (Stephenson 1999).

conference in 2014, Google introduced *material design*, a design language based on the principle of a “material metaphor [as] the unifying theory of a rationalized space and a system of motion”, which is “inspired by the study of paper and ink” (Google 2015a). Material design encourages the use of strong, contrasting colours, large-scale typography and prominent animations (Google 2015a). During a question-and-answer session at the conference, a senior visual designer at Google described material design as being “in a place that’s kind of in between flat and skeuomorphism” (Chiu et al. 2014).

How, precisely, do these design languages differ? What do those differences mean for the designing of apps? How do Android and iPhone users differ, and what do they have in common? Those questions make up the landscape in which this thesis will reside.

1.4 The research problem

Differences between Android and iOS users and the design of a mobile polling application – those are the two main subjects under our consideration. What we are left with, then, are the following two questions –

1. How do iOS and Android users differ in terms of attitude, usage patterns, needs and concerns in the context of a mobile polling application?

2. How to best design the interaction for a mobile polling application for iOS, taking into account the characteristics of the iOS platforms and its users?

1.5 Defining the scope of the thesis

This thesis is concerned with interaction between humans and mobile devices. As such, its scope will not encompass any other sort of interaction, e.g. between humans and personal computers, except as it may relate to mobile interaction.

Moreover, it is specifically concerned with the iOS and Android platforms (even though the FlashPoll project includes a mobile-friendly Web app in addition to the others).

Finally, its scope will not be concerned with other parts of the FlashPoll ecosystem, such as the creation and management of polls, except as it relates directly to the Android or iOS apps.

2

Theory

This chapter will review previous work related to the users of iOS and Android mobiles, the two platforms' design principles, as well as electronic polling, data visualisation and privacy questions. Finally, there is a section on conducting user research.

2.1 Demographics of iOS and Android users

In the year of 2014, for the first time ever, more than a billion smartphones were sold (Gartner 2015). Most of these smartphones run one of two operating systems – Android and iOS. Together these two hold the vast majority of the mobile operating system market – in the second quarter of 2015, Android held 82.8% of the market and iOS 13.9%[1] (IDC 2015). Given this vast market share, the majority of mobile users will own Android or iOS phones, and so it is therefore useful to consider the demographics of these two platforms' users in order to best design an app for them.

The *Pew Research Center* carried out a study in 2013 on smartphone ownership, which, although limited to American adults, sheds some light on the subject (Smith 2013). The study found that some 56% of its participants owned a smartphone. Although the distribution between the Android and iPhone devices was even overall, it differed across various demographic groups – Android devices were adopted evenly across the income and education spectra, whereas persons at the upper end of those spectra were much more likely to own iPhones than were persons on the lower end of the spectra. That is to say both rich and poor, well-educated and not own Android phones, but the rich are more likely than the poor, and the well-educated more likely than those with little education to own iPhones. Moreover, the study found, "African-American cell owners are more likely than whites or Latinos to say that their phone is an Android device as opposed to an iPhone" (Smith 2013). Finally, younger persons were more likely to own Android mobiles, whereas older persons were more likely to own iPhones.

An online dating site looked at users of its Android and iOS apps and found a difference in household income and education levels, where iOS users were more affluent and had higher levels of education than Android users, even though Android users were somewhat older (AYI.com 2014). Moreover, the users of the iOS app

[1] Down 22.3% from the previous quarter in what is a seasonal decline (IDC 2015).

"drank, did drugs, and exercised more often", whereas Android users were more likely to have children and be divorced (AYI.com 2014). These findings, however, need to be taken with more than a grain of salt, as they do not come from a rigorous, scientific study.

How, then, do these two sets of users compare in terms of concerns and attitudes? A study made on user confidence in the privacy and security of mobiles suggests that, while users of Android and iPhone devices were equally concerned with the physical loss of or damage to their phones, Android users were more likely to think about privacy factors when installing new apps than were iOS users, although privacy was "one of the lowest-ranked factors" across users of both platforms (Chin et al. 2012).

This rather slight difference was more pronounced in another study, made with some 700 German students, which found that Android users were more far likely to consider data sharing or permissions when deciding whether to install an app (Benenson et al. 2013). The same study found that those who were brand-aware, i.e. who when choosing a smartphone placed importance in a manufacturer's brand, were more likely to own iPhones than Android mobiles (Benenson et al. 2013).

Felt et al. (2012), notably, found that both the use of location data and the alerting of a user ranked relatively low among respondents' concerns. This study did not, however, mention any differences or likenesses between users of different mobile platforms.

2.2 Android and iOS design principles

This section compares the design principles that were introduced with the flat design of iOS 8 with those that were introduced with the material design of Android 5.0 Lollipop. The following paragraphs will thus be concerned only with the *iOS Human Interface Guidelines* as last updated on December 18, 2014, and with Android's *Material Design – Google design guidelines*, published on November 3 that same year (Apple 2015b; Google 2015a). They will not cover the whole of both guidelines, but rather highlight some key similarities and differences.

2.2.1 Design language

Material design uses paper-inspired *material* – that is, graphical elements with certain common properties of light, shadow, edges, movement etc. – as a way to provide visual cues of the hierarchy and relationship of elements. Actual physical, fibrous paper, then, is the wellspring of material design; the metaphor is one of sheets of material adjacent to or overlapping each other, each of which is holding content.

The flat design of iOS, meanwhile, uses a combination of layering and translucency effects to provide visual cues – views do not change their z-position, so to say, but a sense of depth is nevertheless created through the use of blur effects, more like frosted glass than opaque paper. Instead of grouping and separating content on

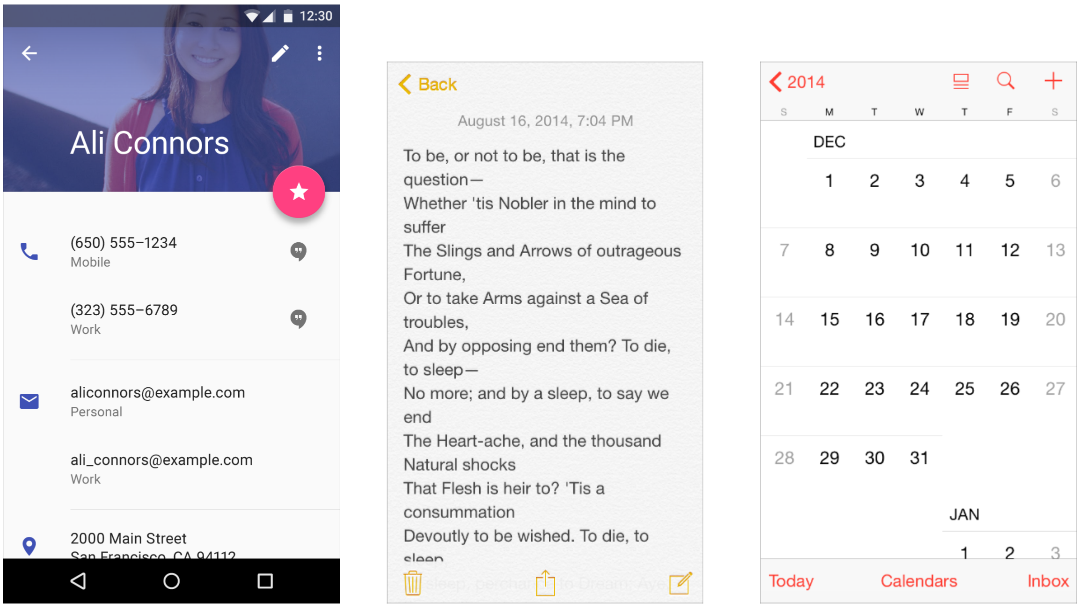


Figure 2.1: Usage of colour in the Android (left) and iOS (centre & right) interfaces. Android's material design emphasises strong, contrasting colours whereas colour in iOS is used more sparingly. Also note the usage of the Roboto typeface (l) and Helvetica Neue (c & r).

rectangular so-called sheets of material, it does so by using negative space (white space). The iOS guidelines declares its three main themes as *deference*, *clarity* and *depth* – that is, letting the interface be subordinate to the content; having sharp, precise and clear typography and graphics; and using the layering and animating of elements to help users' understanding.

A main principle of material design is a focus on strong, contrasting colours, prominent typography and prominent imagery, both as a means to provide meaning and as an aesthetic choice (see Figure 2.1). In the material metaphor, colour is ink, and ink goes on sheets of material, partially or wholly, but never across several sheets. The use of strong, contrasting colours is one of the most noticeable aspects of material design; Google even provides a palette of these strong, contrasting colours, from which the selection of one primary and one secondary colour is suggested, with the secondary, "accent" colour being used mainly to signify important actions.

Unlike material design, the iOS guidelines recommends apps use colour only as "subtle enhancement". Similarly to material design's use of the accent colour, they advise using a key colour to indicate interactivity – in many ways this is the most important function of colour in iOS. When using a colour for this purpose, so Apple notes, one should avoid using it for non-interactive elements also.

As for an app's typography, the two have similar goals but slightly different priorities. Material design uses large-scale typography[2] both as a way of pleasing the eye and

[2]The new version of Android's *Roboto* typeface, which was introduced along with material design, being more similar to iOS's *Helvetica Neue* than its predecessor, is another point of convergence between the two platforms (see Figure 2.1). Apple themselves introduced a new typeface,

to provide hints of meaning and hierarchy. In iOS, the most important function of typography is not its attractiveness but its legibility. Typography in iOS 8 is often smaller than in Android 5.0, although it also lets users adjust the text size across all apps to their liking.

In material design, one groups content by placing it together on a sheet of material. These sheets can be joined or separated, and can overlap each other. While material design makes use of negative space also, in iOS this is one of the main ways of grouping together and separating content.

2.2.2 Graphics and animation

The human interface guidelines of iOS say of an app's user interface that it should not compete with content, but rather help users understand and interact with it. This maxim underpins most of the design principles in iOS, especially as relating to imagery and animation. Visual elements, components and animations are all subtle so as not to distract from the content. This focus is not as explicit in material design, but the emphasis on prominent imagery and typography is still there. The slogan "bold, graphic, intentional" shows how graphics are used in material design: strikingly and unapologetically.

The instructions for designing app icons – the icons that represent the app on the mobile's home screen – are similar in both design languages, and echo the advice found for such icons elsewhere, namely that they should be simple, uncluttered and reflective of an app's brand or identity.

Animation, or motion as it is sometimes referred to in the design documents, is featured prominently in both sets of guidelines. The guidelines emphasise its ability to provide meaning and feedback, and to enhance the user's sense of being able to manipulate things on screen.

However, whereas Apple cautions against overuse of animations, especially ones that seem gratuitous, material design is infused with the usage of animations at all levels, not just to provide meaning (though also that), but also to "delight" the user. Material design, then, prescribes a more liberal, less subtle use of animation than does iOS, including animations for many basic and common actions such as a touch on a button or a touch-and-drag interaction.

2.2.3 Navigation

The first screen that a user sees, so says either set of guidelines, should allow the user to start interacting immediately; that is to say, an app should avoid so-called splash or instruction screens but instead present the user with its core content or functionality. The iOS guidelines categorize the forms of navigation as divided into three styles: hierarchical (e.g. an email client), flat (e.g. a music player) and content-

called *San Francisco*, with iOS 9 (Apple 2015a).

or experience driven (e.g. a game). These styles, of course, may also be combined, and together they do not cover all forms of navigation.

Each platform centres navigation, especially of the hierarchical sort, on a horizontal, top-aligned bar (called navigation bar in iOS and app bar in Android), which contains the current page's title, a button for returning to the previous view and, possibly, buttons for some or other important actions (see Figure 2.2). (Android phones, of course, also have an ever-present "Back" button in addition to the "Home" and "Recents" buttons showing at the bottom of the screen; the physical home button of iOS provides similar functions to the latter two, but navigating upwards in a hierarchy is always done from within the app in iOS.)

For applications with a flat structure, both guidelines recommend the use of a secondary bar (called tab bar or tabs) containing a limited number of buttons for switching between sections within the app. The advantage of this is that the user's navigational options are always visible, and it is clear also in which section he or she currently is. There are a couple of differences here worth mentioning. The tab bar in iOS emphasises icons over texts and is aligned with the bottom of the screen. Android tabs, on the other hand, is located at the top of the screen and contains either icons or, perhaps more commonly, just texts.

One major point of difference between the platforms is Android's recommending the use of a so-called navigation drawer – a vertical menu opened with a button in the top-left corner (see Figure 2.2) – the usage of which iOS strongly discourages. "[The key] things about an intuitive navigation system is that they tell you where you are and they show you where else you can go. Hamburger menus [navigation drawers] are terrible at both of those things", a self-described "user experience evangelist" at Apple has said (Stern, 2014). Indeed, none of the native iOS apps use navigation drawers, whereas Google uses them in their Gmail, Inbox and Docs apps, among others – even as made for iOS.

2.2.4 User interface components

Even more so than the actual guidelines, a platform's native user interface components steer the appearance and the interaction of its apps. Although most of the user interface components found in Android 5.0 and iOS 8 are similar and by now familiar, each of the two platforms have a few distinct such components.

For selection from a set, Android's components are familiar to us from the web – they include checkboxes for choosing multiple options, radio buttons for choosing a single option, and switches for toggling a single option on or off and special pickers for choosing dates and times (see Figure 2.3). iOS, on the other hand, while having a native switch component, lacks both checkboxes and radio buttons. Instead, switches can be substituted for checkboxes when selecting from a set of options, and when it comes to replacing radio buttons, there are several options in iOS, namely the use of either a so-called segmented control – a number of joined, horizontally aligned buttons of which only one can be selected at any time; a picker, which presents a

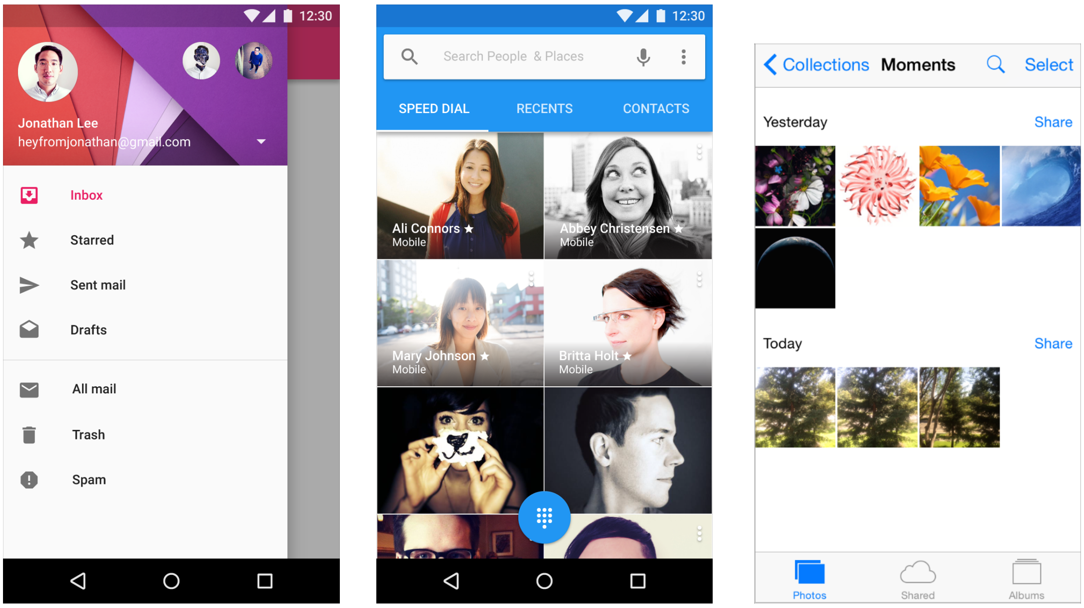


Figure 2.2: Selection of navigation patterns in Android and iOS. Left, the navigational drawer in Android. Centre, tabs in Android. Right, the iOS tab bar.

"spinning wheel" of options, again allowing only one option to be selected; or, if the number of options is very large, a scrollable list.

Both iOS and Android have components for inputting text – single- and multi-line text fields – although the Android versions of these have some additional features (which are described in the guidelines), namely the displaying of feedback such as character count ("42/120") or error messages ("Username or Password is incorrect").

It is worth mentioning, also, what the material design guidelines refer to as the floating action button. Google means for this button, which is circular, elevated and of a bright, eye-catching colour, to be used in signalling actions of especial importance, such as composing a new message in an email client or creating a new file in a file browser (see Figure 2.3). There is no direct equivalent in iOS, where buttons are usually just a coloured icon or verb phrase. In general, buttons and other interactive elements in iOS have relatively low affordances, achieved mainly with colour, placement and context, whereas Android uses a combination of shading and shadows in addition to these means to show that certain elements are interactive.

2.3 Visual information design

Because it deals with polling, the FlashPoll app will naturally deal also with the displaying of quantitative information, more specifically the visual representation of polling results. Cooper et al. (2007) defines visual information design as a field "concerned with the visualization of data, content and navigation rather than interactive functions". Its focus lies in using colour, shape, position and scale to present data in such a way that it's informative and easy to understand (Cooper et al. 2007).

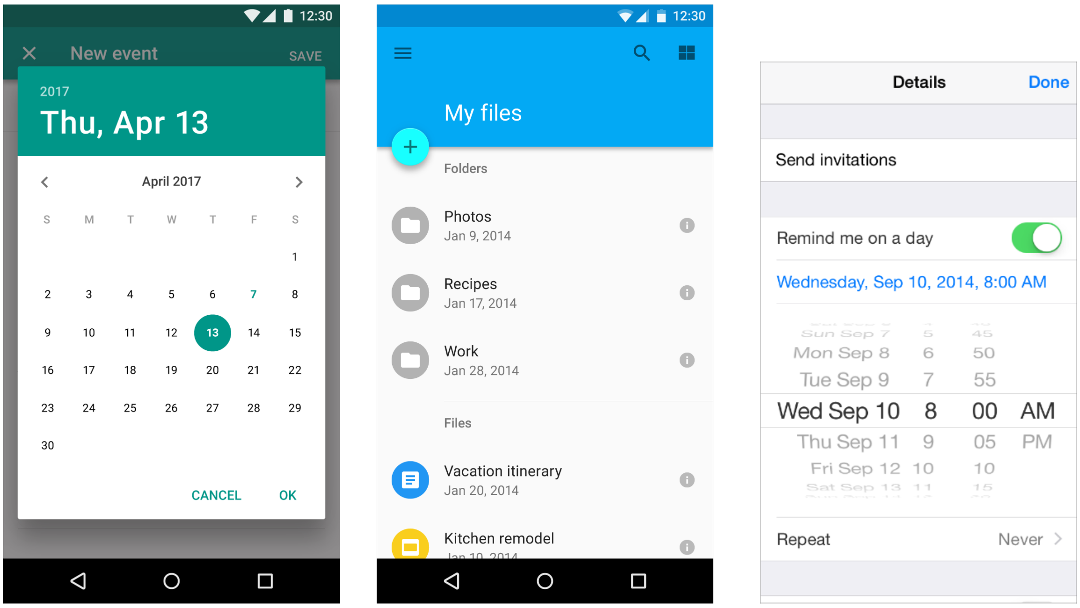


Figure 2.3: Selection of UI components in Android and iOS. Left and right, date pickers in Android and iOS, respectively. Centre, Android's floating action button (the plus sign).

They go on to quote Edward Tufte, writing that good visual design is "clear thinking made visible" (Cooper et al. 2007). In his seminal book *The Visual Display of Quantitative Information*, Tufte (2001) writes about what he terms "data-ink ratio" – the proportion of data conveyed to overall ink usage – and "chartjunk", i.e. generally unnecessary graphical decoration. Distilling Tufte's work into seven principles, Cooper et al. (2007) note that visually displayed information should –

- Enforce visual comparisons
- Show causality
- Show multiple variables
- Integrate text, graphics, and data in one display
- Ensure the quality, relevance, and integrity of the content
- Show things adjacent in space, not stacked in time
- Not de-quantify quantifiable data

The vocabulary for representing data can be extended by the use of animation and interactivity (Ware 2004; Cooper et al. 2007). Motion can be used to express causality (Ware 2004; Michotte 1963) and to gradually transform a structure, e.g. to connect to or separate parts from a whole (Ware 2004). Electronic displays can also allow a user to gradually reveal more detail in the data as he or she wishes (Cooper et al. 2007).

2.4 Designing electronic surveys

Like it did everything else, the Internet changed the nature of polling and surveying a population in a very fundamental way. Web-based surveys, when done well, offer many advantages over print-based ones (Lazar and Preece 1999). However, not much research has been made of them in relation to interaction design, and still less with regard to mobile devices. Still, it is in this context that Nulty (2008) writes: “Suffice it to say that the design of a survey, not only the mode of administration, may also affect who responds to it and what they say.”

Couper (2000) makes the distinction between probability-based and non-probability-based Web surveys. The former are scientific surveys that aim to draw statistical inferences. The latter, which are more germane to our topic, are informal polls, made either for entertainment purposes, like those found on news websites, or as inexpensive, self-selected surveys without any aspirations for scientific validity. He concludes: “We must also learn how to optimally design Web surveys and maximize the benefits of the rich audiovisual and interactive self-administered medium we now have at our disposal” (Couper 2000).

Some research has also been done on how interactivity affects the behaviour of respondents to Web-based surveys, for instance finding that “forced choice formats [used in phone surveys] results in more options being selected” than the “check all that apply” formats which are used in Web-based surveys (Sharp et al. 2007; Smyth et al. 2006). Aday and Cornelius (2006) wrote a book about designing (and conducting) probability-based surveys in the context of health and health care; it touches briefly on Web-based surveys. The authors stress the need for Web-based surveys to be highly accessible and for designers to “design a questionnaire that conforms to what users expect when they operate a computer”. They go on to cite a set of principles given by Dillman (2000), which includes –

- Present questions in a manner similar to that of paper surveys
- Use colour with restraint
- Show very clearly how the user is to proceed
- Don’t force the user to answer before going to the next question

Sharp et al. (2007) also stress the importance of accessibility and advice that, in designing a Web-based questionnaire, it is important to keep privacy and confidentiality questions in mind. Finally, they advice testing the questionnaire thoroughly, for instance using a think-aloud protocol (see Section 3.1.6) (Sharp et al. 2007).

2.5 Privacy

There are two things in the intersection of privacy and interaction design that are of interest to us here: 1) how design affects users’ perception of the risks and benefits associated with privacy issues and 2) how users' privacy perception affects their behaviour.

Privacy-related risk is one of several perceived risks that have been studied in the context of product adoption (Langsner et al. 2006). For instance, Oomen and Leenes (2008) conducted a survey with 5541 Dutch students, finding that participants with higher privacy risk perception did *not* necessarily act more frequently or more strongly to protect their privacy, with a few exceptions such as the use of pseudonyms and the removing of cookies. The same study also found no indication that gender affected privacy risk perception (Oomen and Leenes 2008).

Another study, in which 560 American students were invited to use a geocaching game app once weekly for twelve weeks, found, paradoxically, that its participants disclosed *less* location data as the benefits of doing so increased (Keith et al. 2014). Moreover, they found that participants, when weighing whether or not to disclose information, "based [their decision] on the trustworthiness of the app provider" (Keith et al. 2014).

What, then, of design's impact on users' attitudes regarding privacy? Although the idea that the only purpose of form was to satisfy a material function was prevalent among some of the functionalists in Europe in the 1930s that espoused the principle "form follows function", in reality most were led by a desire for form to *appear* to satisfy function (Hesselgren 1954). A bridge should appear to withstand pressure; a car should look roadworthy; and a ship, seaworthy. To the latter group, what mattered was the set of semantic values that an artistically sensitive designer seeks in the perception of an object (Hesselgren 1954).

Following in their footsteps, designers in the 1980s began to use "product semantics" to inform their designs in order to, as Dunne (2008) puts it, "analyze the way form could be used to express implicit meanings". There are semantic values embedded in the design of interactive objects, some of which values relate to users' perception of privacy. Although modern interactive technologies are in many ways a continuation and digitisation of older technologies, it isn't clear whether, say, perceived risks have simply carried over from older technologies to newer ones, and if so to which extent they have changed in the process. However, one 2003 study did show that the ease of use of an e-commerce site significantly reduced users' risk perceptions in relation to that site (Featherman and Pavlou 2003).

2.6 User research

This section will go over some books and studies on conducting user research. For user research methods, see Section 3.2.

2.6.1 Psychometrics

Psychometrics is the measuring of a person's attitudes, traits, values, opinions, desires, skills or knowledge through various tests. It often takes the form of an attitude scale, the most common of which is perhaps the *Likert scale*, wherein a participant responds to a statement using a certain rating format, usually with 5 or 7 choices, going, for instance, from “strongly disagree” through “undecided” to “strongly agree” (Coolican 2014; Likert 1932). Another common variant is the *semantic differential scale*, which measures a thing’s connotative meaning by asking respondents to place it on a bipolar scale, that is, on a rating format with one descriptor at one pole and another, opposite descriptor at the other pole, e.g. from “bad” to “good” (Coolican 2014; Osgood et al. 1957).

Coolican (2014) identifies three issues co-occurring with psychometric tests: *response acquiescence set* (known also as response bias), *interpretation* and *social desirability*. Response acquiescence set is a phenomenon wherein respondents tend to agree more than they disagree (Coolican 2014). Respondents’ interpretations of questions or statements may affect their emotional state if those questions or statements are extreme or wildly contrary to the respondents’ beliefs; this is likely to cause respondents to answer differently (Coolican 2014). Social desirability means that respondents are more likely to respond with what they believe is socially acceptable, even if it’s not what they actually think – i.e. to choose answers so as to “look good” (Coolican 2014).

2.6.2 Online questionnaires

Electronic surveys are an increasingly used when conducting research (Coolican 2014). Coolican (2014) notes that there are problems involved in gathering a sample population from the Internet, namely that the sample will be skewed (people who use the Internet are not representative of the general population, or there is a self-selection bias involved), although, he notes, “traditional methods [in psychology research] suffer from an over reliance on psychology students”. Moreover, response rates to online surveys are almost always much lower than those for traditional, paper-based surveys (Nulty 2008). One study shows that, although short questionnaires have a higher response rate, “even relatively long questionnaires yield a considerable response rate”, and that “the length of the questionnaire did not have a negative effect on the quality of responses” (Deutskens et al. 2004).

Couper (2000) gives three possible explanations for high nonresponse rates: tried and tested motivating tools (advance letters, personalised signatures) cannot be done in the same way electronically, there may be technical difficulties (browser incompatibility, Internet connectivity issues), and participants may have privacy/confidentiality concerns related to the electronic medium. Aday and Cornelius (2006) propose sending out e-mails with links to the questionnaire as a way of improving the response rate. Deutskens et al. (2004) suggest a lottery with small prizes (but a high chance of winning) as a way to raise the response rate, and, moreover, that this effect is especially important for long questionnaires.

2.6.3 Recruiting participants

In order to make statistical inferences, one should have a probabilistic sample (Couper 2000). That is the ideal – a sample that is a "mini-representation of the population" (Coolican 2014). However, this ideal is likely impossible to attain in practice; instead, one should aim to "remove as much sampling bias as possible" (Coolican 2014).

There are various ways of doing this, both probability-based and non-probability-based, with differing levels of bias involved in each method (Coolican 2014). With Internet samples (samples consisting of participants recruited online) especially, there is typically a self-selection bias, in addition to the inherent coverage bias (Aday and Cornelius 2006).

In the context of quantitative usability studies, Nielsen (2001b) recommends having 20 users for each design being tested in order to get a "reasonably tight confidence interval on the results". Moreover, he reports a 13% no-show rate among user study participants (Nielsen 2003).

3

Methods

This chapter will describe the methods relevant the project – both for the design and development of the app and for the user study. It will also outline some of the methods' respective advantages and disadvantages.

3.1 Design methods

During the course of the project, a variety of methods were considered at various points and for varying purposes. In design, research was done through requirement analyses and heuristic evaluations, and prototyping was used in order to test concepts. Production largely took the form of iterative design, with continuous evaluation using methods such as think-aloud protocols.

What follows is a description of all those methods that were in consideration to be used when designing and making the FlashPoll app.

3.1.1 Requirement analysis

Requirement analysis is a bipartite process: it consists of identifying users' needs and then formalising these needs so as to produce a set of statements about how the product should be or function (Sharp et al. 2007). This set of requirements, which is not static or final, is the foundation on which the design will later be built. Formulating requirements is useful not only later in the design process, but also because it can already at this very early stage reveal problems inherent in the intended product (Sharp et al. 2007).

Cooper et al. (2007) make the distinction between requirements as "needs" and requirements as "features" or "functions" – between focussing on what users actually want versus what engineers or designers *think* they want. Moreover, they propose that thinking the former is superiour to the latter, because it gives the designer greater freedom in future, allowing features and functions to take shape naturally in response to users' needs.

Requirements can also be more easily identified by researching products dealing with similar problems, in this case mainly the FlashPoll app for Android (Sharp et al. 2007), and by considering the preexisting "ecosystem" (including platforms, technologies, standards, technology specifications, etc.) in which the product is to reside. Such reviews familiarises the designer with some of the solutions and limitations found among currently existing products (Cooper et al. 2007).

3.1.2 Heuristic evaluation

Heuristics, e.g. standards or usability guidelines, can be used to evaluate a design cheaply and quickly and without the need to bring in users (Sharp et al. 2007). Perhaps the most well-known set of heuristics, obtained by an empirical analysis, is Nielsen and his colleague's *10 Usability Heuristics for User Interface Design*, which consists of ten useful but non-specific items (Nielsen 1995a). More germane here, however, is Apple's *iOS Human Interface Guidelines*, which exist to ensure a minimum level of usability and consistency in every iOS app. It has advice regarding an app's navigation, graphics, animations, icons and more (Apple 2015b).

3.1.3 Prototyping

Prototypes, the Mus musculi of interaction design, are a means to reflect on a design proposal, to explore ideas and to present them to colleagues or other stakeholders (Sharp et al. 2007). They can represent both large systems as well as small subsets of them, such as specific functionalities or interactions.

The distinction is often made between *low-fidelity* and *high-fidelity* prototypes, the former being simple representations such as paper sketches, mock-ups etc., and the latter being more advanced, often software with relatively complex, interactive elements (Sharp et al. 2007). In a seminal paper, Rettig (1994) recommends low-fidelity prototypes over high-fidelity ones, noting that they require "little more in the way of implementation skills than the ones you learned in kindergarten".

3.1.4 Iterative design

An iterative process is one that allows for reworking parts of a design or product based on new information (Sharp et al. 2007; Cockburn 2007), allowing for the initial design to be imperfect, which is what it's inevitably going to be. As designers and developers engage with the problem at hand, they learn about it and come up with new, improved solutions for it that beg to be realised.

A process which includes abandoning or reversing work already done can, of course, seem Sisyphean, but this point of view misses the fact that the work *did* serve a purpose, that the product would never have reached its current state without having gone through its earlier stages. As Sharp et al. (2007) write, innovation "takes time, evolution, trial and error, and a great deal of patience".

3.1.5 Evaluation in iterative design

As part of the iterative design process, or indeed any design process, one needs to evaluate the ability of a prototype to meet the users' needs. Consider: "[alternative designs] are evaluated through the interactive versions of the designs and the results are fed back into further design" (Sharp et al. 2007). Evaluation is simply a formalised and structured – although not necessarily complex – approach to examining and judging the design of a prototype or product. Sharp et al. (2007) describe three different approaches –

- **Usability testing**, which involves observing or interviewing users as they perform, or after they have performed, a task.
- **Field studies**, which involves evaluating a prototype in an environment similar to that of the product's intended use.
- **Analytical evaluation**, which involves using heuristics or walkthroughs to examine a prototype without the need for users. Heuristics reviews involve using expert knowledge and expertise-based heuristics to examine a system. A cognitive walkthrough entails walking through a task with the aim of noticing usability problems (Sharp et al. 2007).

3.1.6 Think-aloud protocol

The think-aloud protocol is a method wherein a user is asked to report their thought process out loud as he or she does some task or activity (Sharp et al. 2007). It is useful as a way to know what the user is *trying* to accomplish, not just what he or she actually, visibly *is* doing (Sharp et al. 2007). Moreover, it is both cheap and flexible, that is, it can be done with low-fidelity prototypes as well as finished systems (Nielsen 2012).

A problem with the think-aloud method is that it interferes with the usage (having to say everything one does out loud makes cognition more difficult), especially when there is back-and-forth with the person conducting the test (Nielsen 2012). There is also the tendency for a user to lapse into silence after having spoken for a while (Sharp et al. 2007).

3.2 User research methods

This section will describe methods that were researched and considered for the user study.

3.2.1 Usability metrics

Using usability metrics, i.e. quantitative measurements like error rates or time-taking, is expensive, requires more testers and often gives less insight than the alternative, that is to say, qualitative methods (Nielsen 2001b). Why, then, would one wish to measure usability? The main advantage is that it gives precision in research and allows for making comparisons, both between systems and of one system at different points in time (Coolican 2014; Nielsen 2001b). Having numerical data and analysing that data, which is largely what statistics is, is, broadly speaking, an empirical science (Coolican 2014).

In an article on the subject – specifically about user *testing* rather than research – Nielsen (2001b) outlines four basic usability metrics: *success rate*, *time to complete a task*, *error rate* and *users' subjective satisfaction*. The first three of these are also among the nine metrics listed by Wixon and Wilson (1997).

The first and third are simple percentages. The **success rate** is the percentage of tasks that users complete correctly. It is useful mainly because a low success rate is indicative of a serious problem (Nielsen 2001a). Although success is a binary measurement (a user either succeeds or doesn't), it is also possible to include *partial* successes in the metric, giving a more realistic assessment of a system's quality (Nielsen 2001a). The **error rate** can be defined as the percentage of errors made by the user while doing a task. However, unlike successes, errors can be of varying severity – mistyping a message is not the same as a accidentally deleting data (as an example). One way to measure this is to use Nielsen's severity ratings, which, although intended to describe the severity of *problems*, can be adapted for use with errors (Nielsen 1995b). It uses a rating scale going from 0 ("I don't agree that this is a usability problem at all") to 4 ("Usability catastrophe: imperative to fix this before product can be released") (Nielsen 1995b).

The second metric outlined by Nielsen (2001b), the **time to complete a task**, is useful for two reasons: 1) to measure a system's efficiency and 2) to measure its learnability. A task can be expected to take less time when done through an interface that is easier to use, i.e. that is more efficient. At the same time, the time it takes to complete a task is also a measure of learnability because we can expect a task, when repeated several times, to take progressively less time to finish as the user learns about it. In a system that is not learnable at all, the first and the tenth time a task is done will take the same amount of time. Nielsen (2001b) advises using goal-directed tasks that take several minutes to complete, rather than small, quick tasks.

The final metric, that of **users' subjective satisfaction**, simply measures how satisfied users are with a system – how easy they find it to use, how beautiful or ugly it looks to them, and so on. There are several ways of measuring this; we shall consider three methods here – the *System Usability Scale*, the *AttrakDiff* method and *UX Curves*.

3.2.1.1 System Usability Scale

The System Usability Scale (SUS) is a ten-item Likert scale developed by Brooke (1996), aiming to measure a system's overall usability. It consists of statements such as "I think that I would like to use this system frequently" and "I found the system very cumbersome to use" (Brooke 1996). Each item is rated on a 5- or 7-point rating format of the standard Likert sort, for instance "Strongly agree" to "Strongly disagree" (Brooke 1996; Usability.gov 2013). The SUS score is then calculated using the answers to these ten items, to give an overall score on a range from 0 to 100 (Brooke 1996).

The System Usability Scale has been tested empirically and found to be highly robust (Bangor et al. 2008). Efforts have also been made to provide a translation of the scale into German (Lohmann and Schäffer 2013). Besides this, its main advantage lies in its being both short (consisting only of ten items) and easy to understand. However, with this simplicity comes generality – a SUS cannot distinguish, for instance, between a system's visual appeal and its practical usability, nor does it provide any diagnostic information (Usability.gov 2013). Moreover, it only provides a measure of the overall user experience, not of individual parts of such an experience. For instance, if a user is asked to order a book through a Web store, a SUS will then conflate all the individual pages, images, texts and interactions that the user encounters into a single score, not separating between, say, the process of selecting the book and the checkout procedure.

3.2.1.2 AttrakDiff

The AttrakDiff method refers to a 28-item semantic differential scale developed by Hassenzahl (2006), which measures a system's usability in three (or four) qualities, or dimensions. Those qualities are –

- **Pragmatic quality (PQ),** which represents a system's instrumental usability (is it manageable? straightforward? practical?) – the utility it provides the user and the control and confidence with which it allows him or her to use it (Hassenzahl, Kekez, et al. 2002).
- **Hedonic quality (HQ),** which represents its more abstract qualities, those related to the user's *enjoyment* of a system and the pride comes with using or owning it (Hassenzahl, Kekez, et al. 2002). In a revised version of the method, "AttrakDiff 2", Hassenzahl, Burmester, et al. (2003) further divides this quality into two subcategories – **identity (HQ-I)**, which measures how the system relates to the user's identity or self-image (is it integrating? presentable? connective?), and **stimulation (HQ-S)**, which measures the extent to which the system arouses excitement or enthusiasm in the user (is it novel? bold? inventive?).
- **Attractiveness (ATT),** which represents the overall "pull" a system has over the user (is it inviting? attractive? appealing?). A system's attractiveness encompasses its pragmatic and hedonic qualities, that is, there is a positive correlation between the attractiveness and each of these two qualities.

Each of the 28 items is a pair of dichotomous descriptors, such as "bad – good" or "complicated – simple", which the respondent is asked to rate on a 7-point rating format (Hassenzahl, Burmester, et al. 2003; Hassenzahl 2006). There are multiple ways of calculating a score from these responses. A simple way to do it is to calculate the mean value for each quality. This can then either be shown in a line chart or as a "portfolio presentation" showing the system as a rectangle on a two-dimensional grid with the pragmatic quality on the horizontal axis and the hedonic quality on the vertical axis (User Interface Design GmbH 2005).

Like the System Usability Scale, the AttrakDiff method only measures the whole of a user experience. It also does not provide any diagnostic or qualitative information, that is, it doesn't tell you the whys and hows of usability. However, the AttrakDiff method is useful because it separates usability into the three (or four) distinct qualities/dimensions mentioned previously – it distinguishes between a system's practical utility to the user and the "pleasantness" it provides. Finally, it has the modest virtue of originating in the German language: no meaning will be lost in translation.

3.2.1.3 UX Curve

The UX Curve method was developed by Kujala et al. (2011) to provide a tool for measuring users' experiences over time. It does this by having the user draw a curve on a two-dimensional graph, with "the intensity of the users' experience" on its vertical axis and time on its horizontal axis (Kujala et al. 2011). The user is also asked to mark and annotate important changes or events where they occur in the curve (Kujala et al. 2011).

The main strength of UX Curves is, of course, that they, unlike most other methods for evaluating usability, include the temporal aspect of usage, the *long-term user experience*, which can show, for instance, usability improvements (or deteriorations) over time (Kujala et al. 2011). However, the very fact that it *does* entail remembering past events means that it is to a certain degree dependent on the user's recall being accurate, which, as psychology tells us, is not always the case. Indeed, Kujala et al. (2011) acknowledge this, but go on to affirm their belief that "these memories are essential as they are most meaningful to users". The UX Curve method is also slightly cumbersome to collect when compared, for instance, to questionnaires, and is not as easily quantifiable.

3.2.2 Field studies

In order to find out how a product is used in the context *for which it is intended*, by people in their everyday life, it can be useful to conduct a field study (Sharp et al. 2007). There can be said to be two kinds of field studies – those involving *direct observation* (watching participants in person, à la Malinowski) and those involving *indirect observation* (tracking participants from afar). Direct observation offers more control in where to focus attention, but has the disadvantages of being obtrusive and time-consuming (Sharp et al. 2007). Indirect observation, on the other hand, entails observing participants from afar; Bernard (2011) calls it "the archeology of human behaviour". The direct/indirect polarity can be compared with the qualitative/quantitative one, where direct observations often give more qualitative results, and vice versa (Rohrer 2014).

Field studies, which originated in cultural anthropology, have the great advantage of seeing behaviour in its natural context (Coolican 2014). The data that a field study provides is thus contextual and can be both qualitative and quantitative. Inherent in field studies, however, is a certain lack of control as well as a difficulty in recording said data – especially self-reported data can be problematic (Coolican 2014). Sharp et al. (2007) mention some important decisions which need to be taken before undertaking a field study, some of which are: how involved one should be, how to record data and how to handle privacy issues.

3.2.3 Diary studies

In diary studies, participants are given some means (e.g. diary, camera or sound recorder) of setting down their everyday experiences as it relates to a product (Rohrer 2014). "The diary method has the great advantage that the observed persons are acting quite naturally" (Coolican 2014). Diary entries are often made either regularly, often daily, or as triggered by some incident, such as having used the product being studied.

Diary studies are usually longitudinal, making it possible to see changes over time in terms of participants' behaviour, usage or attitudes (Rohrer 2014). They are convenient when participants are geographically dispersed, or when a study is being conducted through the Internet (Sharp et al. 2007). However, there is also the possibility that a participant, when recording an entry in the diary, does not recall very well the day's events, and, moreover, the diary method has, in common with field studies, the dual problems of low control and low participation.

4

Process

This chapter will describe how the work was carried out, from design and development of the app to the carrying out of the user study.

4.1 Design

The major part of the design work done for this project can roughly be said to have been done in three stages – *requirement analysis*, *concept modelling* and *idea generation and prototyping* – all of which took place before development started. The design would undergo changes during development, too, through a process of continuous evaluation and refinement.

4.1.1 Requirement analysis

Before starting the design proper, I made an analysis of the problem domain, looking at literature on the subject, what users of the app would need, what other stakeholders requested, the technical "ecosystem" that would support the app, the existing Android version of the FlashPoll app and other, similar apps, both for web and for mobile. Among other things, the app was required to support iOS 7 in addition to the latest version of iOS, 8.

With all this in mind, I reviewed Apple's *iOS Human Interface Guidelines*, and, having done this, formulated a list of requirements.

4.1.2 Conceptual model

Sharp et al. (2007) define a conceptual model as "an outline of what people can do with a product and what concepts are needed to understand how to interact with it". Johnson and Henderson (2002) characterise it a set of **metaphors/analogies**, **concepts**, **relationships** and **mappings**. Using this template, I noted the following –

- The FlashPoll app's content consists of a collection of *polls*, which, from the user's point of view, are really more like questionnaires or brief surveys.

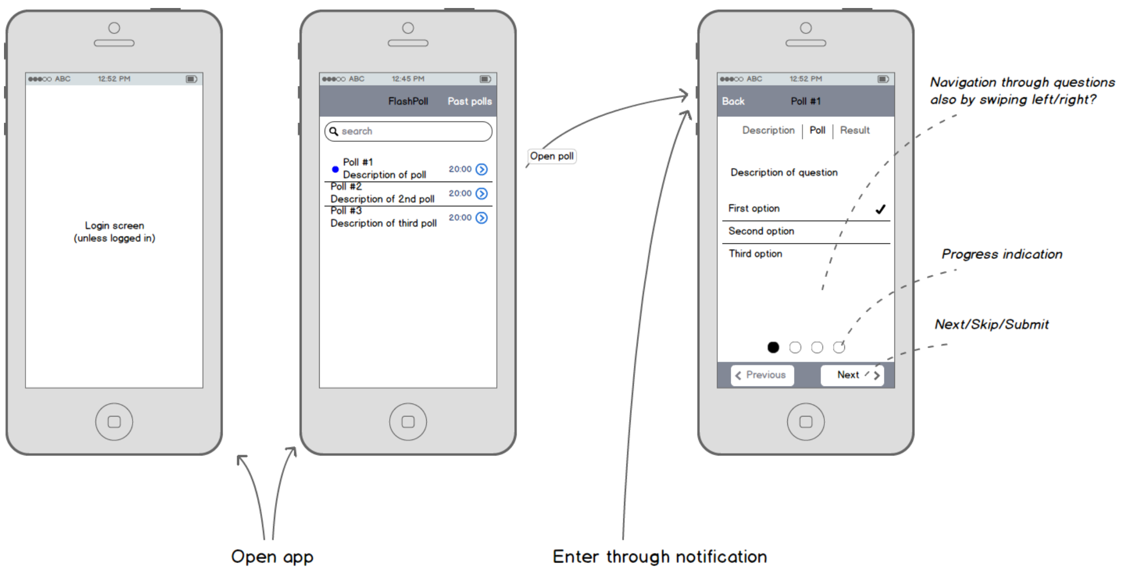


Figure 4.1: Annotated mock-up showing content arrangement and navigational flow.

- A basic item is a *poll*, or questionnaire. Like a questionnaire, a poll has a number of *questions*, each of which question may have one or more *options* available as answers. All these items and subtypes have certain attributes, like titles, descriptions, texts and so on.
- A poll consists of one or more questions. Each question may have any number of answer options.
- Each poll in the system corresponds to one answerable survey for the user.

4.1.3 Prototyping

Because the FlashPoll iOS app had a rather tight deadline, the initial design process had to be relatively compressed in order that development could begin soon after. Still, it obviously being important to arrive at an initial concept even in an agile process, I spent some time on prototyping to try out different ideas.

Based on the requirements, and using the conceptual model, I made sketches on paper exploring how content could be arranged, navigation organised and information visualised effectively. These were followed by an annotated mock-up (see Figure 4.1) which, although simple, covered the major part of the system, including its content hierarchy and navigational flow. I bounced this off stakeholders for feedback and to make sure it was in line with their requests and users' needs.

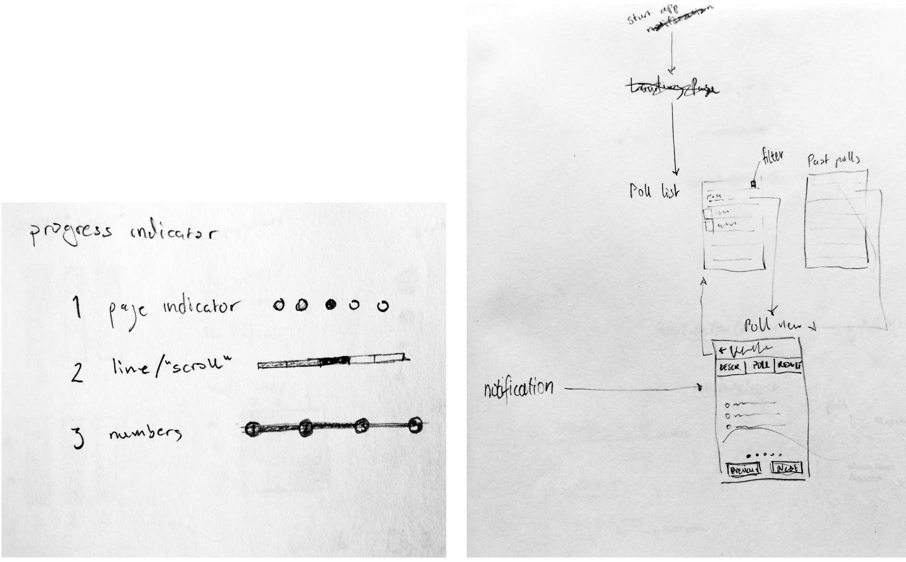


Figure 4.2: Excerpts from the design notebook. Left, concept sketches for the progress indicator, showing which question (or result) the user is currently viewing in a poll. Right, sketch of navigational structure.

4.1.4 Sketching

Throughout the project, I put down ideas, drawings and concepts in a design notebook (see Figure 4.2). This served both as a way to remember ideas and concepts and to present those ideas and concepts to others. For instance, before prototyping or implementing the results visualisation for ranking questions, I sketched several proposals and used the sketches as basis for a brainstorming session. This lead both to a refining of the proposals and a narrowing down of proposals worth pursuing.

Moreover, I used the notebook to take notes during meetings and to organise lists of things that needed doing, although most of the planning and scheduling was done digitally.

4.2 Development

Having thus made an initial design concept, I began the process of developing the app, while continuously evaluating what I had made and making changes and improvements to the design. The app was developed with *Xcode* (v6.3 and v6.4), Apple's standard workspace for making iOS and OS X applications. *Git* (v2.3.6) was used for version control, which allowed for reversal to previous states of the code, dividing work into different branches and easily sharing code and documentation with colleagues and stakeholders.

In early iterations, the app was made to show dummy polls, without any of the underlying logic, i.e. it was not integrated with the FlashPoll ecosystem but more resembled a high-fidelity prototype. The very first iteration of the app, which included the whole navigational structure as well as much of the UI functionality in a

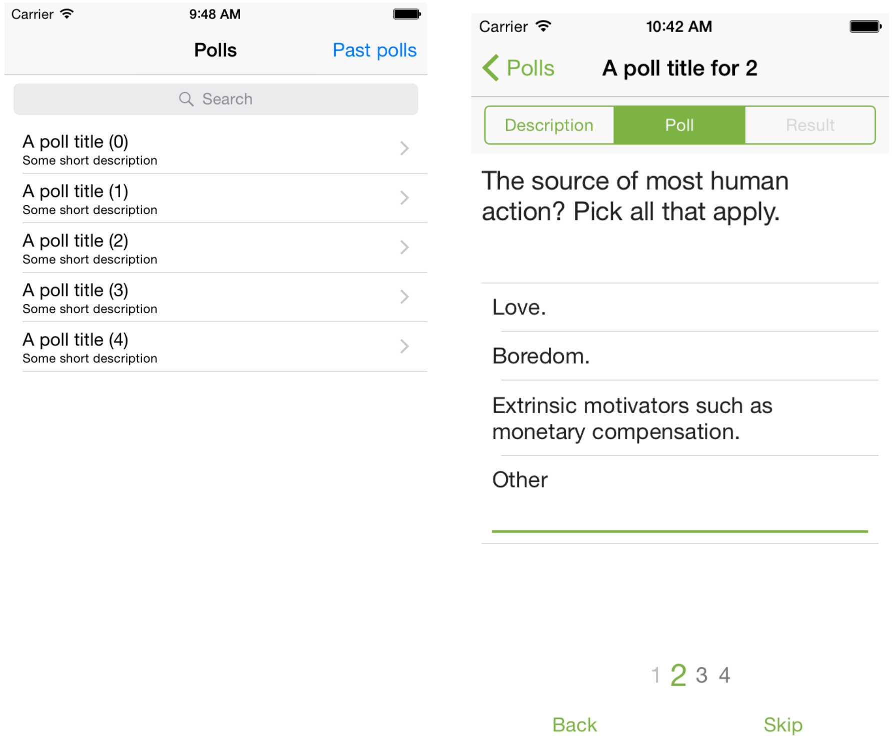


Figure 4.3: Screenshots of the app in very early iterations. Left, the poll list view as it looked two days into development. Right, the poll view, 11 days into development.

working state, was up and running on the first day, allowing for evaluation from a very early stage. Figure 4.3 shows two screenshots of the app in its early stages.

4.2.1 Continuous evaluation

Because the app was always kept in a working, if unfinished, state, I was always able evaluate it, and then based on an evaluation refine its design and implement changes. This process was continuous and cyclical, and went on until very late in the development.

Evaluation was mainly carried out in two ways (see Section 3.1.5 for details on evaluation methods): through simple, informal **usability tests** and through **analytical evaluations** such as cognitive walkthroughs. The former was mainly done by observing persons as they performed a task with the app, sometimes using a think-aloud protocol. For instance, a user could be asked to respond to a certain poll, to delete a poll or to find a poll by searching for it. This was followed by a brief and informal interview. Testers could range from friends or family members who were previously unfamiliar with the app to various stakeholders who knew it well.

The other method used was cognitive walkthroughs. This entailed performing the same tasks as those mentioned in the previous paragraph, i.e. responding to a poll, deleting a poll etc., and then using knowledge about the subject and the domain, as well as various usability heuristics, chiefly the *iOS Human Interface Guidelines*, to identify both the elements in the app that were confusing or problematic as well as those that worked well.

This process of continuous evaluation resulted in several changes being made to the design, including but not limited to:

- Initially, the app divided “active polls” (i.e. polls that the user can respond to) and “past polls” (i.e. polls that have expired) into separate lists, allowing the user to switch between the two with a button in the top-right corner of the view (see Figure 4.3). This had a couple of disadvantages: 1) it was guaranteed to hide a subset of the polls at all times and 2) it involved extra tapping on the user’s part when switching between the two.

  A better solution was deemed to have both active and past polls in the same list, grouped under separate headers. The app wasn’t likely to have that many active polls available at any one time, and even if there were, it is very easy to scroll in a list on an iPhone. Moreover, the view was decluttered by removing the button in the title bar, which also allowed for moving the search field up, freeing up space for what was important: the poll list.
- Noticing that it was not obvious how, having opened a poll, to proceed to a poll’s actual questions, a “Start” button was added.
- Text labels displaying the expiration dates of polls (i.e. the date on which it is no longer possible to respond to a poll) was changed from showing dates, e.g. “12 May, 2015”, to showing the actual remaining time, e.g. “5 days” or “2 hrs”. This was done because the date itself was mostly irrelevant to users – what he or she wanted to know was how much time remained until he or she could no longer respond to the poll.
- Originally, the design was identical for both multiple-choice and single-choice questions, with selected options being marked with only checkmarks. This was changed to the more common representations of checkmarks (for multiple-choice) and circles (for single-choice). New icons were also made for options that weren’t selected, allowing users to easily see which type of question it is, even when no option is selected.

4.2.2 Technical limitations

Some circumstances impact a product’s design even if they have no intrinsic relation to interaction design. For instance, a product’s design may be impacted by certain technical complications or time constraints. The changes that these circumstances necessitate are generally unwanted and, unlike the changes listed above, do not arise out of an effort to remedy issues found in the app. Here are two such changes made to the FlashPoll app:

- Originally, the app was made to work both in portrait and landscape mode. However, in evaluating the app, I found that it was almost never the case that portrait mode wasn't the superiour way of present information; in fact, that for every view, landscape mode was unnecessary. Because of this, the app was made to be locked in portrait mode.

- Relatively late in the project, it emerged that the underlying data structure for a poll's result (decided upon before this project began) did not include detailed results for questions of the "ranking" type. That is to say, information about how many persons had selected each option was unavailable; instead, the system assigned a score to each position (e.g. 3, 2 and 1 for a question with three options) and summed up all the scores for each option, so that, for each option, only a weighed score was available. This necessitated changes in visualising the results of ranking questions.

4.3 User study

Planning for the user study began as development reached a stage where the app was relatively finished. Planning was done on and off for several weeks, the study itself lasted around two weeks and analysis of the gathered data went on during many weeks after that.

4.3.1 Research questions

Looking back at the research problem, we want to know how Android and iPhone users differ in terms of "attitude, usage patterns, needs and concerns". In addition to this – both in order to gauge the relative significance of the findings and for its own sake – it may be of interest to look at differences between other groups, too, such as gender, employment status and level of education. Given this, I defined six general research questions for the user study –

- How much do participants use their mobiles?
- What do they use their mobiles for?
- How much do they use the FlashPoll app?
- When do they use the app?
- Where or in which context do they use the app?
- What were their reasons for using it or not using it?

For each of these questions, there were two added questions: 1) whether there are any **differences between groups**, particularly between Android and iPhone users, and 2) whether there are any **correlations with participants' attitudes or traits**. For instance, if Android users tended to use the app during the day and iPhone users during the night, or if there was a correlation between privacy concern and time spent using the app, that was something we wished to know.

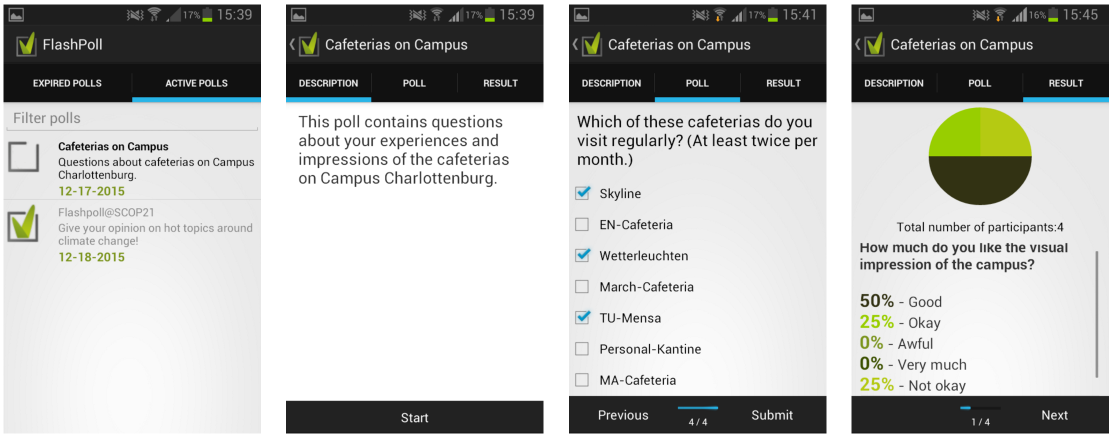


Figure 4.4: Screenshots of the Android version of the FlashPoll app, which was made about a year prior to the iOS version. Far left, the list of polls. Centre left, the poll description, which is shown when a poll is opened. Centre right, answering a multiple-choice question. Far right, result for a single-choice question. (For screenshots of the iOS version, see Section 6.2.)

4.3.2 Planning the study

How, then, to best find answers to these questions? Because the FlashPoll app is location-based, it did not make sense to test it only in a controlled, geographically static setting. A field study, on the other hand, which would allow participants to use the app in its natural setting, seemed like the better choice. This field study would include daily *diary reports* by the participants, allowing me to examine how usage changes over a week. My supervisor at the Quality and Usability Lab deemed 40 participants to be enough, about half of whom would be Android users and half iPhone users; testing 20 users per design is also what Nielsen (2001b) recommends.

Having decided on the format, I drafted questions for the diary reports and for a final, *end-of-survey questionnaire*, which would include all those questions that only needed asking once, such as demographical information and attitudes on issues such as privacy and civic engagement. These questions went through several iterations, changing both as I reviewed them and as I got advice in meetings with my supervisor.

In trying to compare and contrast Android and iPhone users, and especially their respective behaviours and usage patterns, it was necessary to first evaluate and compare the Android and iPhone versions of the FlashPoll app. If one of the two was, for instance, more user-friendly than the other, that was likely to affect participants' usage of the apps. To solve this problem, the apps' respective user interfaces would need to be tested under controlled conditions. For this test, I decided to have a subset of the participants (40 being excessive) perform two tasks with each of the apps and then report their opinions. The tasks would consist of the most common use cases: responding to a poll and looking at the results for a poll.

Of Nielsen's four basic usability metrics (see Section 3.2.1), three were relevant here: error rate, task completion time and users' subjective satisfaction (the fourth metric, success rate, was irrelevant since the tasks are simple enough that participants will always succeed). This meant that, while performing the tasks, participants would be observed in order to identify common mistakes made while using the app *(error rate)*, and timed in order to get a measure of the interfaces' efficiency *(task completion time)*. To understand how participants experienced the apps *(users' subjective satisfaction)*, I considered evaluation questionnaires such as the *System Usability Scale*, the *AttrakDiff* method and *UX curves* (see Section 3.2), eventually deciding on AttrakDiff, since it had been used previously at the Quality and Usability Lab and allowed more fine-grained insight (separating, for instance, between pragmatic and hedonic qualities). Think-aloud protocols were rejected because they would affect time measures.

Three reasons called for meeting with participants in person. First, the aforementioned user interface evaluation would benefit greatly from being conducted in person, allowing for observations, accurate timekeeping and the ability to clear up misunderstandings in testing. Second, it was clear from previous studies in which the FlashPoll app for Android had been used, which had been conducted *entirely* remotely, that problems were inevitably going to arise as soon as participants were asked to install the app on their own devices. To make things worse, the installation procedure for the iPhone version of the app was more complicated and comprised more steps than that of the Android version. Third, participants had to sign consent forms in order to take part in the study. For these reasons, it was deemed best to have participants come in for a "starting session" before the beginning of the field study.

Finally, I wrote a protocol for the starting session, outlining point-by-point each step that would make up the session. Moreover, I wrote an informational leaflet to give to participants and an advertisement to recruit them. A colleague translated this along with all of the relevant polls, questionnaires and other information into German.

4.3.2.1 Preparing the questionnaires

As mentioned, the end-of-survey questionnaire would contain questions not only about demographical information, but also to measure various attitudes and skills. In order to do these measurements, we used various Likert scales taken from previous research (see Table 4.1). All of the scales were used unchanged (albeit translated into German) apart from the "Proficiency with mobiles" scale, which was adapted to the usage of mobile devices specifically. Because both "AttrakDiff" and "Big-Five personalities" were originally defined as 7-point scales, all the other scales were made to use the 7-point format, too, making the questionnaires more consistent overall. Nearly all of the scales had a mixture of positive and negative statements, so as

[1] Used for the user interface evaluation and not in the end-of-survey questionnaire. Split into 4 subscales: pragmatic quality, hedonic quality (identity), hedonic quality (stimulation) and attractiveness.

[2] Split into 5 subscales: extraversion, agreeableness, conscientiousness, emotional stability and openness to experiences.

Attitude scale used in the study	Number of items	Source
AttrakDiff[1]	28	Hassenzahl, Kekez, et al. (2002)
Big-Five personality test[2]	10	Gosling et al. (2003)
Proficiency with mobiles	6	Fogarty et al. (1999)
Civic engagement	8	Doolittle and Faul (2013)
Mobile users' information privacy concern	21	Malhotra et al. (2004) and Xu et al. (2012)
Normative privacy beliefs	12	Unpublished study
Desire for privacy	6	Morton (2013)
Perceived risks with sharing location information	13	Unpublished study
Perceived benefits with sharing location information	7	Unpublished study

Table 4.1: Table of attitude scales used in the user study, showing the number of items included in each scale and the source from which the scale was taken. (The AttrakDiff method actually uses a semantic differential scale, but is included here for convenience.)

to avoid response acquiescence (a bias in which respondents tend to agree with all statements).

The diary reports and the end-of-survey questionnaire were added to *LimeSurvey*[3], which is the standard tool used for this purpose at the Quality and Usability Lab. The visual appearance of Likert scales on LimeSurvey (see Figure 4.8), along with the actual wording of choices, also enhanced the notion that options were equidistant, meaning that results could be treated as interval-level data rather than ordinal later on.

4.3.2.2 Preparing the apps

In order to get a fuller understanding of how participants use the apps during the field study – which naturally does not take place in a controlled setting – we set up a system in both the Android and iOS app where they would send messages to a server whenever the app was being used[4], which server would then store the information for later analysis. Messages included "opened app", "opened poll", "viewed question", "submitted response to poll" and so on. I also set up a (local) web page (see Figure 4.5) displaying a list of all these messages.

Moreover, a special version of the app was made for the user interface evaluation,

[3]An open-source application used for administering online surveys (Schmitz 2015).

[4]For privacy reasons, participants were made aware that usage data would be collected prior to taking part in the study. Data was never collected outside the confines of the user study, such as from the version on the App Store.

Usage data

Device id	Time	Type	Poll id	Section id	Question id	Skipped	Response
FPUsageStudy_USR_01	7/13/2015, 9:38:23 AM	openedPoll	6dd06b99-0047-4c13-958c-c3ac05a54e91				false
FPUsageStudy_USR_01	7/13/2015, 9:38:31 AM	changedPollSection	6dd06b99-0047-4c13-958c-c3ac05a54e91	1			false
FPUsageStudy_USR_01	7/13/2015, 9:39:54 AM	closedPoll	6dd06b99-0047-4c13-958c-c3ac05a54e91				false
FPUsageStudy_USR_01	7/13/2015, 9:40:33 AM	openedPoll	6dd06b99-0047-4c13-958c-c3ac05a54e91				false
FPUsageStudy_USR_01	7/13/2015, 9:40:35 AM	changedPollSection	6dd06b99-0047-4c13-958c-c3ac05a54e91	1			false
FPUsageStudy_USR_01	7/13/2015, 9:40:38 AM	answeredQuestion	6dd06b99-0047-4c13-958c-c3ac05a54e91		0		false
FPUsageStudy_USR_01	7/13/2015, 9:40:41 AM	viewedQuestion	6dd06b99-0047-4c13-958c-c3ac05a54e91		1	false	false
10726549827424524614	7/13/2015, 9:40:41 AM	openedApp					false
10726549827424524614	7/13/2015, 9:40:41 AM	closedApp					false
10726549827424524614	7/13/2015, 9:40:49 AM	openedApp					false
FPUsageStudy_USR_01	7/13/2015, 9:40:55 AM	answeredQuestion	6dd06b99-0047-4c13-958c-c3ac05a54e91		1		false
FPUsageStudy_USR_01	7/13/2015, 9:40:56 AM	viewedQuestion	6dd06b99-0047-4c13-958c-c3ac05a54e91		2	false	false
FPUsageStudy_USR_01	7/13/2015, 9:41:09 AM	answeredQuestion	6dd06b99-0047-4c13-958c-c3ac05a54e91		2		false
FPUsageStudy_USR_01	7/13/2015, 9:41:13 AM	viewedQuestion	6dd06b99-0047-4c13-958c-c3ac05a54e91		3	false	false
FPUsageStudy_USR_01	7/13/2015, 9:41:24 AM	closedPoll	6dd06b99-0047-4c13-958c-c3ac05a54e91				false
10726549827424524614	7/13/2015, 9:41:43 AM	closedApp					false
RobDevTest	7/13/2015, 9:47:47 AM	openedPoll	370bc3f9-9bf6-45f3-819e-4fa7324b8da2				false
RobDevTest	7/13/2015, 9:47:52 AM	changedPollSection	370bc3f9-9bf6-45f3-819e-4fa7324b8da2	1			false
RobDevTest	7/13/2015, 9:47:54 AM	changedPollSection	370bc3f9-9bf6-45f3-819e-4fa7324b8da2	0			false
RobDevTest	7/13/2015, 9:47:55 AM	changedPollSection	370bc3f9-9bf6-45f3-819e-4fa7324b8da2	1			false
RobDevTest	7/13/2015, 9:47:56 AM	changedPollSection	370bc3f9-9bf6-45f3-819e-4fa7324b8da2	0			false
RobDevTest	7/13/2015, 9:47:57 AM	closedPoll	370bc3f9-9bf6-45f3-819e-4fa7324b8da2				false
RobDevTest	7/13/2015, 9:48:05 AM	closedApp					false
RobDevTest	7/13/2015, 9:56:31 AM	openedApp					false
FPUsageStudy_USR_02	7/13/2015, 10:21:57 AM	closedApp					false
FPUsageStudy_USR_02	7/13/2015, 10:22:29 AM	openedApp					false
FPUsageStudy_USR_02	7/13/2015, 10:22:30 AM	closedApp					false
RobDevTest	7/13/2015, 10:23:33 AM	openedPoll	370bc3f9-9bf6-45f3-819e-4fa7324b8da2				false
RobDevTest	7/13/2015, 10:23:37 AM	pressedStartPoll	370bc3f9-9bf6-45f3-819e-4fa7324b8da2				false
RobDevTest	7/13/2015, 10:23:37 AM	changedPollSection	370bc3f9-9bf6-45f3-819e-4fa7324b8da2	1			false
RobDevTest	7/13/2015, 10:23:41 AM	answeredQuestion	370bc3f9-9bf6-45f3-819e-4fa7324b8da2		0		false
RobDevTest	7/13/2015, 10:23:42 AM	viewedQuestion	370bc3f9-9bf6-45f3-819e-4fa7324b8da2		1	false	false
RobDevTest	7/13/2015, 10:23:57 AM	viewedQuestion	370bc3f9-9bf6-45f3-819e-4fa7324b8da2		2	false	false
RobDevTest	7/13/2015, 10:24:02 AM	answeredQuestion	370bc3f9-9bf6-45f3-819e-4fa7324b8da2		2		false
RobDevTest	7/13/2015, 10:24:04 AM	answeredQuestion	370bc3f9-9bf6-45f3-819e-4fa7324b8da2		2		false
RobDevTest	7/13/2015, 10:24:10 AM	viewedQuestion	370bc3f9-9bf6-45f3-819e-4fa7324b8da2		3	false	false
RobDevTest	7/13/2015, 10:24:18 AM	answeredQuestion	370bc3f9-9bf6-45f3-819e-4fa7324b8da2		3		false
RobDevTest	7/13/2015, 10:24:20 AM	answeredQuestion	370bc3f9-9bf6-45f3-819e-4fa7324b8da2		3		false
RobDevTest	7/13/2015, 10:24:22 AM	answeredQuestion	370bc3f9-9bf6-45f3-819e-4fa7324b8da2		3		false

Figure 4.5: Web page showing usage data recorded from the apps.

which version contained a predetermined set of polls (to ensure consistency between tests) and allowed for resetting the state of the app, so that a poll could be responded to several times.

4.3.3 Pilot study

Having outlined the study and prepared the necessary materials, I conducted a pilot with 3 participants – 1 Android user and 2 iPhone users. The pilot started with a session on the 2nd of July and continued for two days. Apart from the smaller number of participants and the shorter time frame, as well as the smaller number of polls available, the pilot imitated the real study closely, using the same material, the same apps and the same protocol.

During the pilot study, while carrying out the user interface evaluations, almost all of the errors made were caused by "minor usability problems", which meant that Nielsen's severity ratings (see Section 3.2.1) did not add much value, but, on the contrary, made it more difficult to effectively track mistakes, since judging a problem's level of severity took attention away from the participant's actions. For this reason, I decided to simply count mistakes that occurred during usage, not classifying or labelling them.

Moreover, it became clear that the method for distributing the iOS app to participants' mobiles – using Apple's *TestFlight* service – was saddled with a couple of drawbacks: 1) it did not support iOS 7 (only iOS 8 and later) and 2) it did not allow for changes to be made to the app without going through an approval process, which process could sometimes last a week or more. For this reason, I made the decision to switch to the *Crashlytics* service, which *did* support iOS 7 and which *would* allow updates to be distributed to participants instantaneously, without going through an approval process. The main drawback of Crashlytics, meanwhile – namely a more complicated installation procedure – was manageable, since installation was done on location with the researchers' assistance.

More generally, there was some confusion among participants about the study, how the field study would work, what was required of them and how they should respond to the surveys. To rectify this, I changed the protocol to include exactly what each participant needed be told, especially regarding how the study would proceed after the opening session.

4.3.4 Recruitment and participants

As the study was focussed on comparing Android and iPhone users, I used the sampled by quota in order to ensure roughly half of participants were Android users and half iPhone users. Once the quotas were filled, because of time and budget constraints, I sampled by convenience, albeit with an attempt to obtain as diverse (in gender, age and employment status) a sample as possible.

Participants were recruited mainly through four channels – an advertisement on Prometei (a platform maintained by TU Berlin and used for finding participants), an advertisement on eBay.de, word of mouth, and flyers put up on campus and in the city. To keep the drop-out rate as low as possible, participants were offered a reward: upon completion of the study, a participant would receive 8-15 € depending on their diligence in filling out their diary entries (15 € minus 1 € for each missed diary report).

In the end, most of the participants, 27 out of 40, were recruited through the Prometei platform. A handful were friends or acquaintances. The rest responded to the eBay advertisement or to the flyers.

4.3.5 Carrying out the study

The study was carried out between the 13th and 27th of July with 40 participants, of whom 18 were owners of iPhones (45%) and 22 of Android phones (55%).

Having agreed to take part in the study, participants were asked to schedule an appointment for the starting session. They could choose from time slots spread over three days' time, with starting times varying from 9 in the morning to 8 in the evening, depending on the day. Most time slots allowed for up to 3 participants to sign up at a time, although some slots – those that were reserved for participants who would also take part in the user interface evaluation – were limited to a single participant.

The main part of the study took place in the course of ten mostly sunny summer days in the city of Berlin. The starting sessions ran for three days (Monday through Wednesday) and were followed by the week-long field study (ending on the following

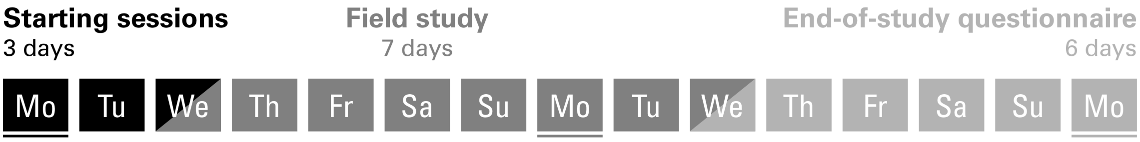


Figure 4.6: Timeline of the user study, including the starting sessions (during which the user interface evaluation took place) and the field study, wherein participants used the app and submitted daily diary reports. The end-of-study questionnaire was administered online and took about 20 minutes to complete.

Figure 4.7: The location of the user study's starting sessions.

Wednesday (see Figure 4.6), although participants were given some additional time after that date to fill out and submit the final online survey).

The sessions took place in the privacy of a light-controlled and soundproofed room on the 17th floor of the Telefunken-Hochhaus (see Figure 4.7). They were conducted by myself and a German-speaking colleague. On arriving, participants were introduced to the FlashPoll app and to the format of the study. They were also told that comments and answers, if they end up published, will not be attributable to them; and that what is important in the study is how good the *information* provided is, not how well or often the app is used.

Having been briefed on the study, participants were asked to sign a consent form stating that they have received information about the study and that their data will be stored and used for scientific purposes. They were then assisted in getting the app installed on their phone. Finally, some users also took part in the user interface evaluation.

The whole of the study – including oral instructions, handouts and other materials, surveys, questions and email correspondence, as well as the apps and polls themselves – was conducted in the German language.

4.3.5.1 Evaluating the user interfaces

In order to evaluate the two apps' user interfaces, a small study was made in conjunction with the main study, using 10 of the 40 participants. This sampling was made in order that of those 10, 5 would be iPhone users and 5 Android users. Apart from that, as well as scheduling considerations, I aimed for as diverse (in gender, age and employment status) a sample as possible.

The evaluation was made with one participant at a time. It started with the participant being given instructions on how the evaluation would proceed: being first asked to perform two simple tasks with one of the two apps, and then to fill out a brief questionnaire for that app before doing the same procedure with the other app. The tests of the two apps were separated in this way (as opposed to performing all tasks first before filling out any of the questionnaires) in order not to have the participant confuse the two apps or be encouraged to compare them explicitly.

Before performing the tasks, we told the participant that we were testing the app, not the participant. We also told the participant that he or she did not need spend much time thinking about the answers while carrying out the tasks (since what is important is not the answers but the *act* of answering), and that we would be observing him or her and taking notes as he or she used the app.

4.3.5.2 Studying the apps in the field

Before the study began, 55 polls were created with varying subjects, lengths, expiration dates and geographical distributions (see Appendix 2 for examples). For instance, some polls concerned the city of Berlin and thus covered its whole geographical area; others concerned neighbourhoods, parks, libraries, malls, cinemas and so on, thus covering smaller areas. Some were set to expire during the study; others at the end of it. All of them were made to be as realistic as was feasible within the constraints of the study.

These polls were set to start at some point during the field study (usually at its outset). The exception to this was one poll, which was set to start before the study began, in order to have a poll available for participants the moment they installed the app.

Each participant was asked to use the app as he or she may normally – as much or as little as he or she wants to (or even not at all) – for a week's time (see Figure 4.6). After each day, the participants were asked to fill out a diary report – a brief online survey asking them about how much they used their phone, what they have used it for and why, whether and in which context they did or did not use the FlashPoll app. The survey was made available at 16:00 each day and lasted until 15:59 the following day, with reminders being sent out to those participants who had yet to answer some 4-5 hours prior to the survey's expiring.

During this field study, usage data was logged directly from the devices (see Section 4.3.2.2), which data included events such as the participant opening or closing the

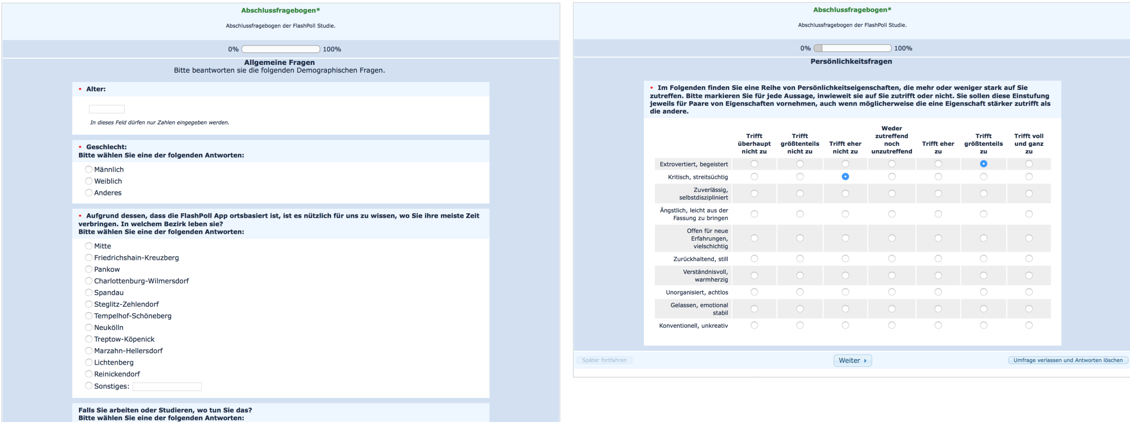


Figure 4.8: Screenshots of the end-of-study survey, which was administered online. The diary reports were done in the same fashion.

app, initiating a search, navigating through the app, answering a question and submitting a response.

After the week had passed, the participants were asked to fill out the end-of-study questionnaire, containing questions about demographical information and Likert scales aimed at measuring various attitudes and skills (see Figure 4.8). This questionnaire ended with a debriefing, wherein participants were informed of the polls used during the study's having been created by the researchers (not by any of the purported organisations or agencies), and that this had been done in order to create as realistic and lifelike an environment as possible for the study.

Upon submitting their response to the end-of-study survey, participants were thanked for having taken part in the study and sent their rewards in the form of Amazon gift vouchers.

4.3.5.3 Interpreting the data

Once the the study had ended, I made an analysis of all of the collected data, statistics and answers. Data from the questionnaires and directly from the apps was compiled, organised and formatted using scripts written in the *Ruby* programming language (v2.2.1p85). This organised and formatted data was then imported into *RStudio* (v0.99.441), where calculations were made and graphs drawn in the *R* language (v3.2.0).

In order to estimate the reliability of the Likert scales that were administered – both the AttrakDiff questionnaire and those in the end-of-study survey – Cronbach's alphas were calculated[5] for each of them. If a scale was found to have low reliability, an item analysis was made; if the analysis showed that deleting an item would significantly increase the alpha, that item was then deleted. This procedure was repeated until a good alpha was reached or until the number of items was low enough that the scale was no longer valid.

[5]These scale analyses were done in Ruby, using the `statsample` (v2.0.1) library.

The data was then analysed by performing various calculations:

- **Calculating means and standard deviations.** Participant age is one example of a variable which was satisfactorily explained by means and standard deviations. The AttrakDiff method (see Section 3.2.1.2) also uses these values for evaluation.
- **Looking for differences between two different groups.** The Mann-Whitney U test was used when looking at differences between two groups, such as between Android and iPhone users or between genders.
- **Looking for differences between three or more groups.** ANOVA was used in order to detect differences between three or more groups, such as between participants with different levels of education or employment status. If there was a significant difference between groups, post-hoc analysis was made through pairwise Mann-Whitney U tests with Holm-Bonferroni correction.
- **Looking for correlations between two variables.** The Pearson's r measure was calculated when looking for correlations between two variables, such as between a participant's level of civic engagement and the total time he or she spent with his or her phone.

5

Results

This chapter will summarise the results of the thesis work: it will describe app itself in its final state, present the results of the subsequent user study and, finally, propose a number of guidelines for similar designs problems in future.

5.1 The FlashPoll application

5.1.1 The list of polls

The app, as it looks in its final state, can be said to consist of two views: the list of polls and a single poll. The first, the *poll list view* (see Figure 5.1), is what the user is presented with after opening the app: it shows a list of all the polls which the user has received – displaying a name, description and closing date for each item – grouped into two categories, “Active Polls” (at the top) and “Expired Polls” (below). A checkmark appears next to a poll’s title and description in the list if the user has responded to that poll.

A search field, embedded in the navigation bar, allows the user to search for polls by name or description (see the centre left image in Figure 5.1). Polls can be removed from the list using the standard iOS interaction of dragging an item leftward and then pressing the “Delete” button that appears (see the centre right image in Figure 5.1). The user can also use a pull-to-refresh gesture (as is a standard idiom in iOS) in order to manually make the app fetch any newly available polls (see the far right image in Figure 5.1).

Because the app is reliant on a number of user-dependent conditions in order to function – it needs to be able to access the Internet, to track the user’s location and to receive remote notifications – it may need to present the user with messages about these things. For instance, it may ask the user to “Give the FlashPoll app permission to receive notifications so that it can find [the user] polls”. These messages are shown above the poll list, as can be seen in the image on the left in Figure 5.2, and may allow the user to tap on them in order to open the Settings app, where the user can make the necessary change (see the centre image in Figure 5.2).

If the app is running in the background when a new poll becomes available (e.g. by the user entering the poll’s geographical area), a notification appears on the phone

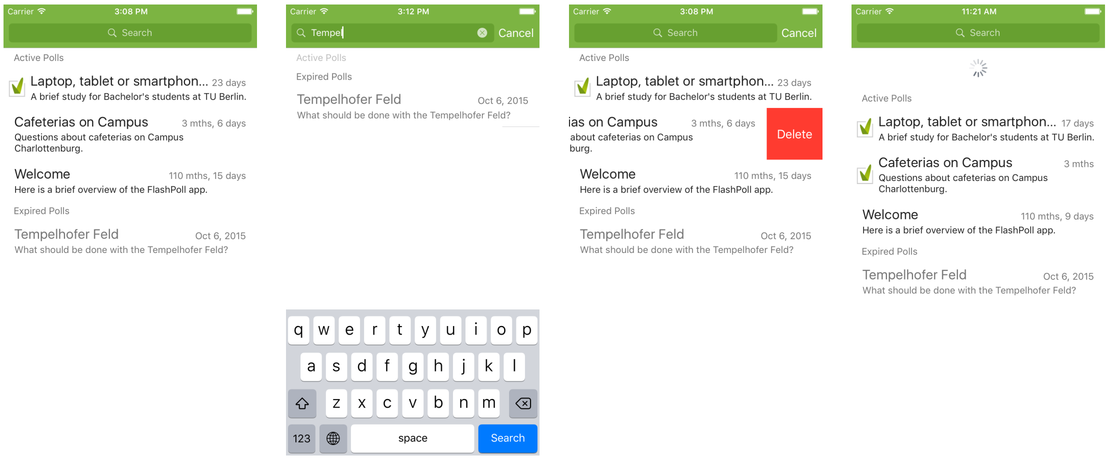


Figure 5.1: Screenshots of the poll list view in its final state. Far left, the list of polls as it appears after opening the app. Centre left, searching in the list. Centre right, removing a poll. Far right, the pull-to-refresh indicator.

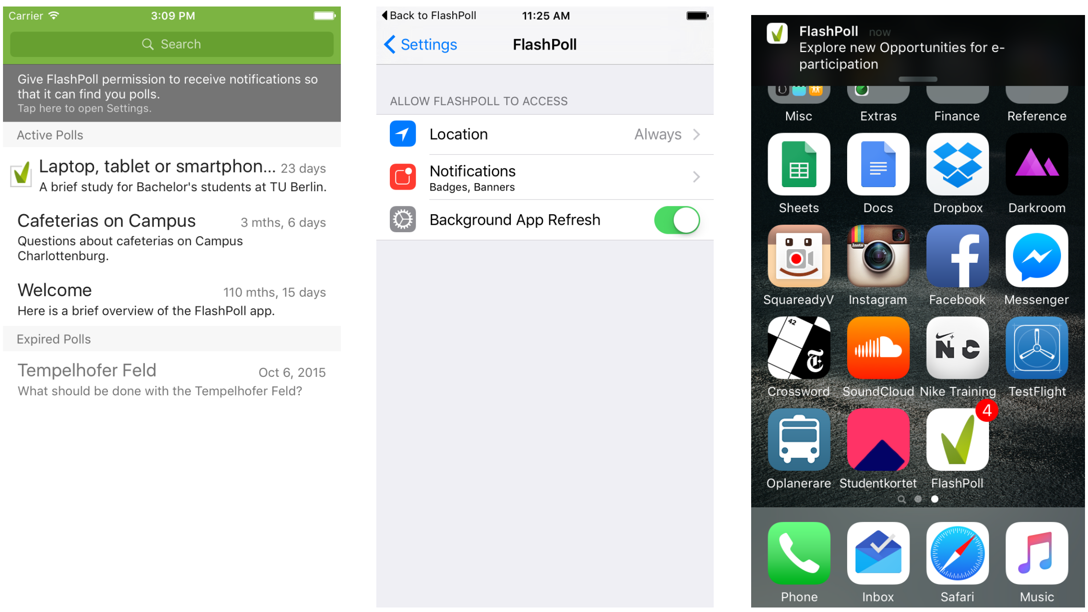


Figure 5.2: Screenshots of the app in its final state. Left, a message asking the user to give the app permission to receive notifications. Centre, the app's section in the Settings app. Right, receiving a poll notification.

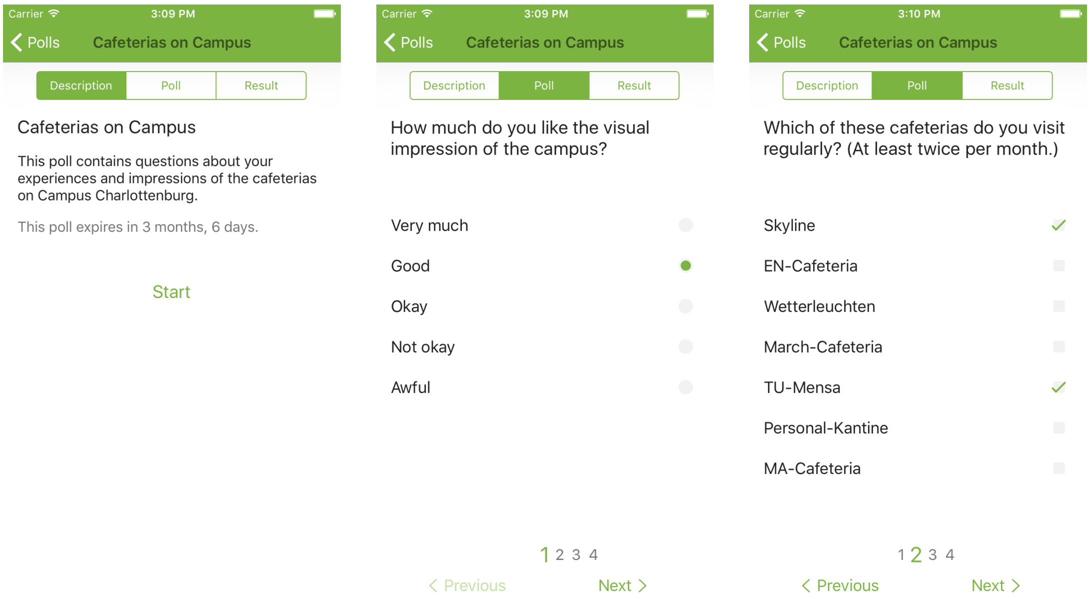


Figure 5.3: Screenshots of the poll view in its final state. Left, the poll description, which is shown when a poll is opened. Centre, answering a single-choice question. Right, answering a multiple-choice question.

(see the image on the right in Figure 5.2). The app's badge – the red circle at the upper right corner of the app icon – is also updated to indicate the number of new polls.

5.1.2 The poll view

When the user taps on an item in the list, they are presented with the *poll view* (see Figures 5.3, 5.4 and 5.5). This view consists of three sections, or subviews, each of which can be accessed from a menu at the top: *Description*, *Poll* and *Result*. Navigating between sections sets off a brief and subtle animation where the current view "disappears into" the current section, and the new view "springs out of" the chosen section – this is done to strengthen the association between each section's button and view. At any time, the user can return to the poll list by tapping the "Polls" button in the navigation bar (as is standard behaviour in iOS).

The Description section (see the image on the left in Figure 5.3) presents some general information to the user: the poll's title, a long description and its closing date (in a longer format). A "Start" button brings the user to the Poll section.

The Poll section is where the user responds to a poll. It consists simply of a series of questions, each of which has a description text and an input fields of varying type – radio buttons for single-choice questions, checkboxes for multiple-choice questions, re-arrangable (via touch-and-drag) options for ranking questions and text fields for free text questions (see Figures 5.3 and 5.4 for examples of these views).

The user can navigate between the questions either by scrolling left and right or by using the "Next" and "Previous" buttons. If a question is mandatory, the "Next"

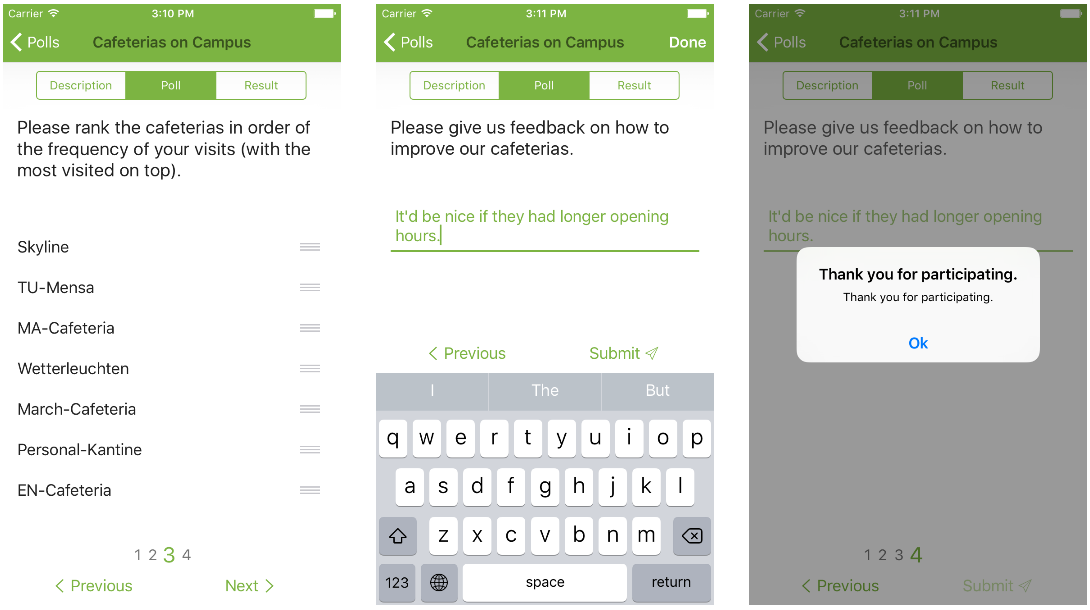


Figure 5.4: Screenshots of the poll view in its final state. Left, answering a ranking question. Centre, answering a free text question. Right, alert shown after having submitted one's response.

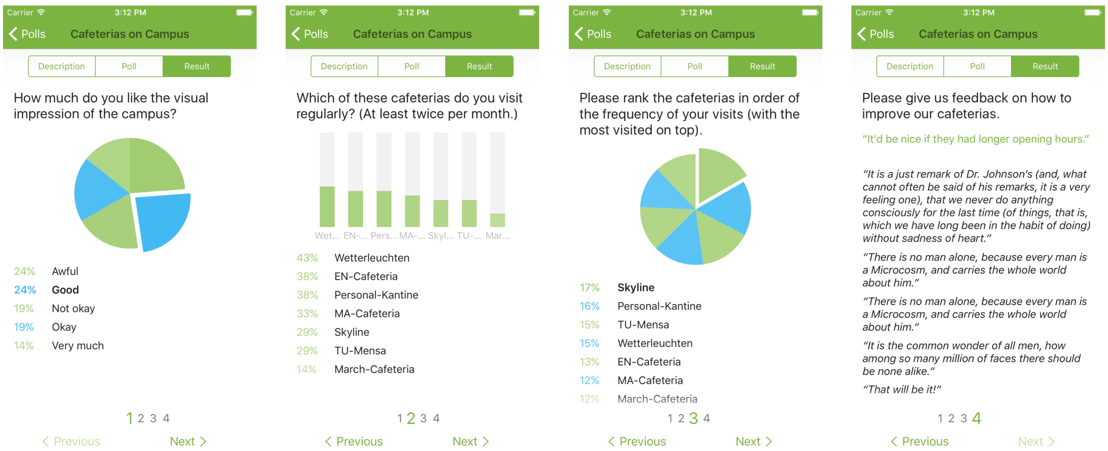


Figure 5.5: Screenshots of the poll view in its final state. Far left, result for a single-choice question. Centre left, result for a free text question. Centre right, result for a multiple-choice question. Far right, result for a ranking question.

button is disabled until an answer has been provided. When the user has reached the final question, the "Next" button becomes a "Submit" button. After having submitted a response to a poll, the user is shown an alert (see the image on the right in Figure 5.4), and then (having tapped "Ok") redirected either to the Result section (if the result is set to be visible) or the poll list (if not).

The Result section is similar to the Poll view: like the Poll view, it consists of a series of questions navigable either through scrolling or using the "Next" and "Previous" buttons. The only difference is that, for each question, instead of an input field, it shows a visualisation of the result for that question. See the images in Figure 5.5 for examples of these views.

For single-choice questions, results are shown in the form of a pie chart, as well as all the options listed by percentages; individual slices can be selected and highlighted by tapping on them, or by tapping on the item in the list below. For multiple-choice questions, this visualisation is replaced by a bar chart. Results for ranking questions also use a pie chart for visualisation; here, the percentages refer to an aggregate[1] of ranking placements for each item. Finally, results for free text questions are visualised simply by showing first the user's own answer (in green) and then the five most recent answers.

5.2 The user study

The user study, after having been carried out (see Section 4.3), resulted in a trove of data, both qualitative and quantitative. That data was then analysed[2]; what follows is a summary of that analysis.

5.2.1 Validity

All 40 participants (see Table 5.1 for the sample's demographic characteristics) completed the field study and submitted their answers to the end-of-study survey. Of the 280 occasions for diary entries (one per day per participant, that is, 40 participants times 7 days), 12 were missed, meaning that 95.7% of all diary entries were submitted as expected. 31 participants (77.5%) submitted all seven of their diary entries, and only 2 participants missed more than one entry.

Results related to one participant, who selected the same option for all of the statements in the end-of-study survey (rather than actually responding to the questions), were excluded from all calculations except those related solely to the user interface evaluation, where the user's data was valid.

[1]For instance, if options A, B and C are ranked in that order in a response, their scores would be increased by 3, 2 and 1, respectively. The total score for all options is the aggregate of all responses.

[2]Performed in the R language (v3.2.0) using RStudio (v0.99.441). Additional preprocessing of data was done in Ruby (v2.2.1p85).

Demographic	Percentage of participants (%)
Gender[3]	
Male	58.97
Female	41.03
Mobile platform	
Android	56.41
iPhone	43.59
Mobile subscription with Internet access	
Yes	89.74
No	10.26
Education[4]	
Abgeschlossenes Studium (Uni, FH)	43.59
Abitur (inkl. Fachabitur)	35.90
Mittlere Reife	20.51
Hauptschulabschluss	0
No degree	0
Employment	
Student	51.28
Employed full-time or more	23.08
Employed part time (less than 35 hours per week)	10.26
Self-employed	7.69
Unemployed	7.69
Retired or disabled	0

Table 5.1: Table showing the demographic characteristics of the sample. $N = 39$.

Additionally, there were two cases for which participants' daily estimated times spent with their phones were ignored. First, when they had entered wildly improbable numbers (spending, for instance, as much as 48 hours with their phone in a single day). Second, when they had entered the same number for all or all but one of the days, e.g. spending exactly 500 minutes with their phone each day. This was the case for the results of 3 participants in total. However, the results of those participant were used where this data was not relevant, for the reason that their other responses were deemed probable and credible.

[3] There were no participants of non-binary gender.

[4] There are no direct international equivalents of these degrees, but they can be roughly considered to mean the following (ordered by distinction). *Hauptschulabschluss* = certificate given for completing Hauptschule (secondary school). *Mittlere Reife* = certificate given for completing Realschule (secondary school) and passing a final exam (comparable to American high school diploma). *Abitur* = certificate given for completing Gymnasium (secondary school) and passing a final exam, opening access to higher education. *Abgeschlossenes Studium* = university degree (bachelor's or master's).

Likert scale	Standardized Cronbach's α
Civic engagement	0.8967
Perceived location-sharing benefits	0.8872
AttrakDiff	
HQ-S	0.8871
ATT	0.8762
HQ-I[5]	0.8173
PQ[6]	0.8104
Mobile users' information privacy concern	0.8527
Proficiency with mobiles	0.8133
Perceived location-sharing risks	0.8061
Normative privacy beliefs	0.7615
Desire for privacy[7]	0.7087

Table 5.2: Table of standardized Cronbach's α for the Likert scales used in the study. The scales' internal consistency range from very good (~0.9) to acceptable (~0.7). (The AttrakDiff method actually uses a semantic differential scale, but is included here for convenience.)

Participants had a mean age of 29.5 years (31.5 years for Android users and 27.0 years for iPhone users; $N = 39$). Together, they covered ten of the twelve boroughs of Berlin, with most living in Mitte (20.5%) and Charlottenburg-Wilmersdorf (17.9%). As can be seen in Table 5.1, a little over half of the sample was made up of students; there was also a tendency among participants to have high levels of formal education.

The user interface evaluation was carried out with a subsample of 10 of the 40 participants, of which 5 were Android users and 5 iPhone users. One of the participants in this subsample was female and 9 male; 2 were unemployed, 3 employed (full-time, part-time or self-employed) and 5 students. They had a mean age of 26.3 years.

Finally, during the study, it was discovered that a few of the Android participants had been prompted to update the FlashPoll app to the version on Google Play, thus stopping those participants from seeing the polls made for the study. As soon as this discovery was made, emails were sent out to the affected participants asking them to reinstall the version that was made for the study. The impact this is likely to have caused is very low.

5.2.1.1 Reliability of psychometrics

Generally, a Cronbach's α of 0.75 or more is considered to represent good reliability (Coolican 2014). As can be seen in Table 5.2, the internal consistency of the scales used in this study range from acceptable ("Desire for privacy" had an α of around 0.7) to very good ("Civic engagement" had an α just below 0.9).

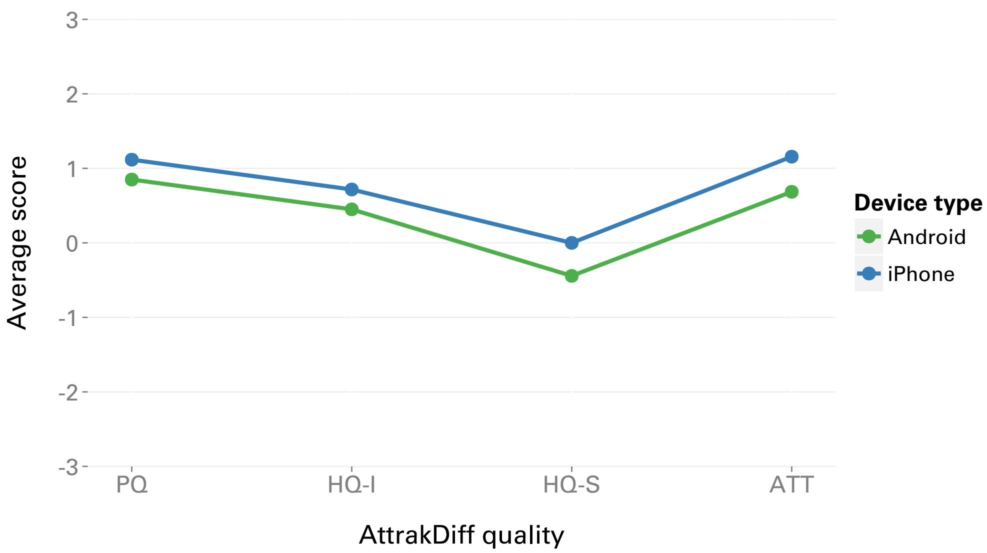


Figure 5.6: AttrakDiff scores for the two versions of the FlashPoll app (separated into *pragmatic quality*, *hedonic quality [identity]*, *hedonic quality [stimulation]* and *attractiveness*). $N = 10$.

In order to increase the reliability of the "Desire for privacy" test, which initially had a Cronbach's α of 0.6633, I made an item analysis and deleted its final item ("I try to change the topic of a conversation if people start asking too much about me."), which improved the scale's α to a more acceptable value: 0.7087. This same procedure was done with two qualities of the AttrakDiff scale: the pragmatic quality ("PQ"), which had a new α of 0.8104 after deleting "unprofessional – professional", and the hedonic quality (identity) ("HQ-I"), which had a new α of 0.8173 after deleting "unpredictable – predictable".

5.2.2 User interface evaluations

The two versions of the FlashPoll apps (Android and iOS) were evaluated using the AttrakDiff method (see Section 3.2.1.2 for an explanation of AttrakDiff). Results show that the iOS version was slightly better received than the Android version in all four qualities (see Figure 5.6 and Table 5.3). Participants commented that the Android version was "a little colourless" and that, when the keyboard was visible, there was "too much on the screen"; however, they also wrote that it was "clearly arranged" and had a "nice presentation". For the iOS version, comments said that it "looked nice" and that its "layout in general was pretty pleasant", but also noted that, with regard to the sorting of the options for ranking questions, it "wasn't

[5] The second item in this scale ("unpredictable – predictable") was deleted in order to improve the scale's reliability.

[6] The fifth item in this scale ("unprofessional – professional") was deleted in order to improve the scale's reliability.

[7] The last item in this scale ("I try to change the topic of a conversation if people start asking too much about me") was deleted in order to improve the scale's reliability.

Quality	Android		iPhone	
	μ	σ	μ	σ
Pragmatic quality	0.85	1.06	1.12	0.80
Hedonic quality (identity)	0.45	1.09	0.72	0.99
Hedonic quality (stimulation)	-0.44	1.10	0	0.85
Attractiveness	0.69	0.72	1.16	0.95

Table 5.3: Results of the AttrakDiff questionnaire in the user interface evaluation. See Figure 5.6 for a graphical view.

Measurement	Android		iPhone	
	μ	σ	μ	σ
Mistakes	1.00	1.00	1.60	1.14
Time to complete first task (s)	70.26	14.53	78.21	28.70
Time to complete second task (s)	31.19	14.52	26.03	13.30

Table 5.4: Measurements of errors made and the time taken to complete the tasks in the user interface evaluation.

obvious on the first view how it works".

As can be seen in Table 5.4, participants made more errors with the iOS version ($\mu_i = 1.60, \sigma_i = 1.14$) than with the Android version ($\mu_a = 1.00, \sigma_a = 1.00$). Notes show that nearly all of these errors occurred when answering the first question of the "ranking" type – most participants struggled at first with understanding the drag-and-drop interaction (however, all participants eventually figured it out). "Sorting is not intuitive enough", one participant wrote about the iOS version.

Finally, we look at the times taken to complete each of the two tasks: answering a poll (T_1), and viewing a certain result for a poll (T_2). The mean time to complete the first task was $\mu_1 = 74.23$ seconds ($\sigma_1 = 21.85$), whereas the mean time for the second task was $\mu_2 = 28.61$ seconds ($\sigma_2 = 13.40$). Participants took slightly longer to complete the first task with the iOS version than the Android version (see Table 5.4), but slightly less time to complete the second task. Overall, the differences between the two apps were not major. We can also note that the standard deviation was high for the iOS version's time with the first task, where some, though not all, participants had difficulties with ranking questions.

5.2.3 Demographic findings

Before looking at the app usage, it is worth considering the make-up of various subsets of the sample. In terms of personality, proficiency with mobiles, privacy attitudes and civic engagement, there were no significant differences between Android-

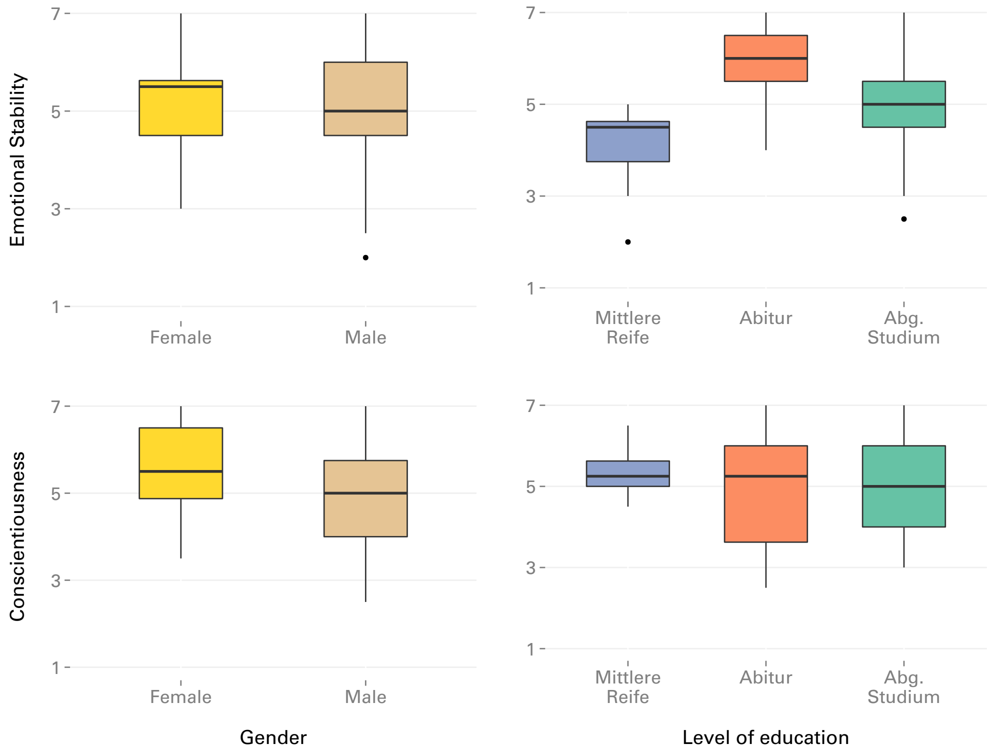


Figure 5.7: Box plots showing differences between participants with different genders (left) and levels of education (right). The vertical axes show participants' *Emotional Stability* (top) and their *Conscientiousness* (bottom) as drawn from the Big-Five personalities measure. Note the differences in conscientiousness between genders (bottom-left) and in emotional stability between participants with Mittlere Reife and Abitur educations (top-right). $N = 39$.

and iPhone-using participants or between participants with different employment statuses. However, there *were* significant differences in terms of gender and level of education, specifically in terms of *Conscientiousness* and *Emotional Stability* (see Figure 5.7).

First, there was a significant difference (Mann-Whitney $U = 253, N_f = 16, N_m = 23, p = .049$ two-tailed) in Conscientiousness between female and male participants, with means of 5.56 and 4.74 respectively (see the bottom-left plot in Figure 5.7). In other words, female participants were significantly more conscientious than were male participants.

Second, a one-way ANOVA showed a significant difference ($F(2, 36) = 7.930, p = .001$) in Emotional Stability between participants with different educations. A post-hoc test using Mann-Whitney tests with Holm-Bonferroni correction showed that *participants with an Abitur education had higher emotional stability than those with a lower level of education (Mittlere Reife)* ($p = .005, r = .144$; means were 5.82 and 4.06 respectively; see the top-right plot in Figure 5.7). No significant differences were found either between participants with Abgeschlossenes Studium and Abitur educations ($p = .056$) or Abgeschlossenes Studium and Mittlere Reife ($p = .056$).

5.2.4 Examining the research questions

Turning to look at the actual results, we recall the questions set out before the user study's beginning –

- How much do participants use their mobiles?
- What do they use their mobiles for?
- How much do they use the FlashPoll app?
- When do they use the app?
- Where or in which context do they use the app?
- What were their reasons for using it or not using it?

This section will deal with each of these questions in order.

5.2.4.1 How much did participants use their mobiles?

As we can see in Figure 5.8, the time that participants used their phones (*not* only the time that they spent with the FlashPoll app), as estimated by the participants themselves, was relatively stable over the course of the week. The overall median time that participants spent with their phones each day was 100 minutes (1h 40m); the mean was 137.7 minutes (2h 18m) and the standard deviation 103.7. Notable, perhaps, is the temporary dip in time spent with the phone which was reported on Sunday, i.e. for the period between Saturday afternoon and Sunday afternoon.

No significant differences in overall phone usage were found between Android and iPhone users, nor by gender, level of education or employment status.

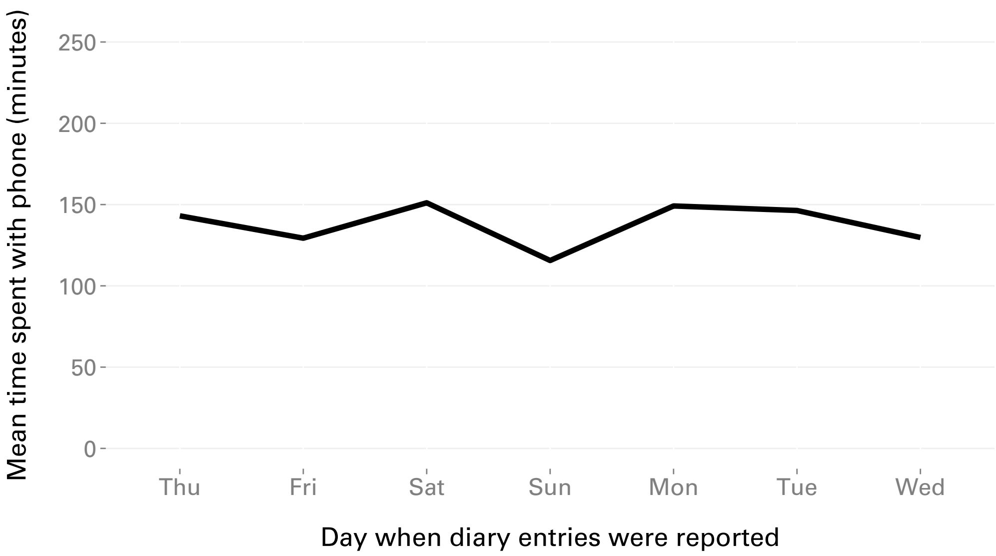


Figure 5.8: Mean time per day that participants used their phones, in their own estimation. The days on the horizontal axis signify the days on which the diary entry was submitted (and so concern a 24-hour period beginning on the previous day's afternoon). $N = 36$.

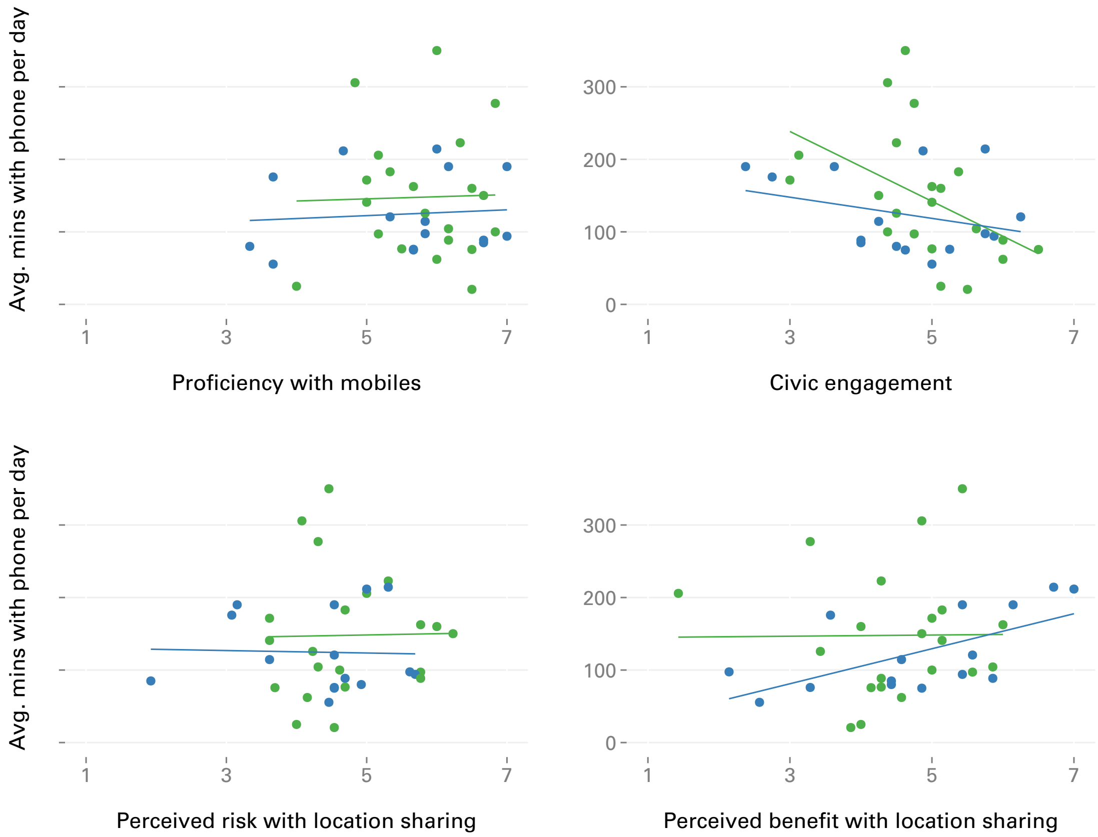


Figure 5.9: Scatter plots illustrating relationships between how much time participants used their phones each day and various psychometric measurements. Each dot is a participant; Android users are green and iPhone users blue. $N_a = 21, N_i = 15$.

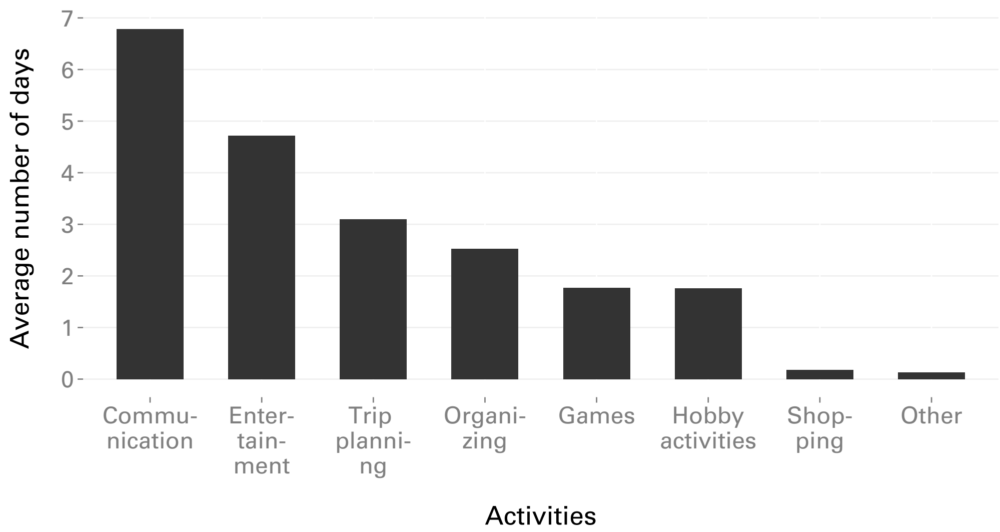


Figure 5.10: Bar chart showing what participants used their phones for – the vertical axis shows the average number of days on which participants reported using the phone for a certain activity, from 0 (not on any days) to 7 (on every day in the study). $N = 39$.

Two notes, however. First, as can be seen in Figure 5.9, there was a strong, negative correlation ($r(19)$ = -.490, p = .024) between Android users' *civic engagement* and the time they spent with their phones. That is, for Android users, having a stronger civic engagement was correlated with using one's phone *less* every day. This correlation was not significant for iPhone users.

Second, as is also shown in Figure 5.9, there was a strong, positive correlation ($r(13)$ = .633, p = .011) between iPhone users' *perceived benefits* of sharing location information and the time they spent with their phones. That is, for iPhone users, seeing more benefits in the sharing of private information was correlated with using one's phone more every day. However, no such correlation was found among Android users, and, moreover, no correlation was found in either group between perceived *risks* and the time spent using their phones.

5.2.4.2 What did they use it for?

In Figure 5.10, we can see that participants used their phones for *communication* (calling, texting, messaging) nearly every day: it had a mean value of 6.782 days[8]. This was followed by *entertainment* (browsing the web, watching videos; $\mu = 4.72$), *trip planning* (booking flights, finding public transport connections; $\mu = 3.10$) and *organizing* (managing finances, tasks or calendars; $\mu = 2.53$).

[8]These numbers are normalised such that 100% usage equals 7 days, even if the number of submitted diary entries is lower. That is to say, if, for instance, a participant has submitted 6 diary entries and reported as having used their phone for "Communication" thrice, i.e. on 3 of those entries, that would count as $\frac{3}{6} \cdot 7 = 3.5$ normalised days, that is, half of the full week considered in the study.

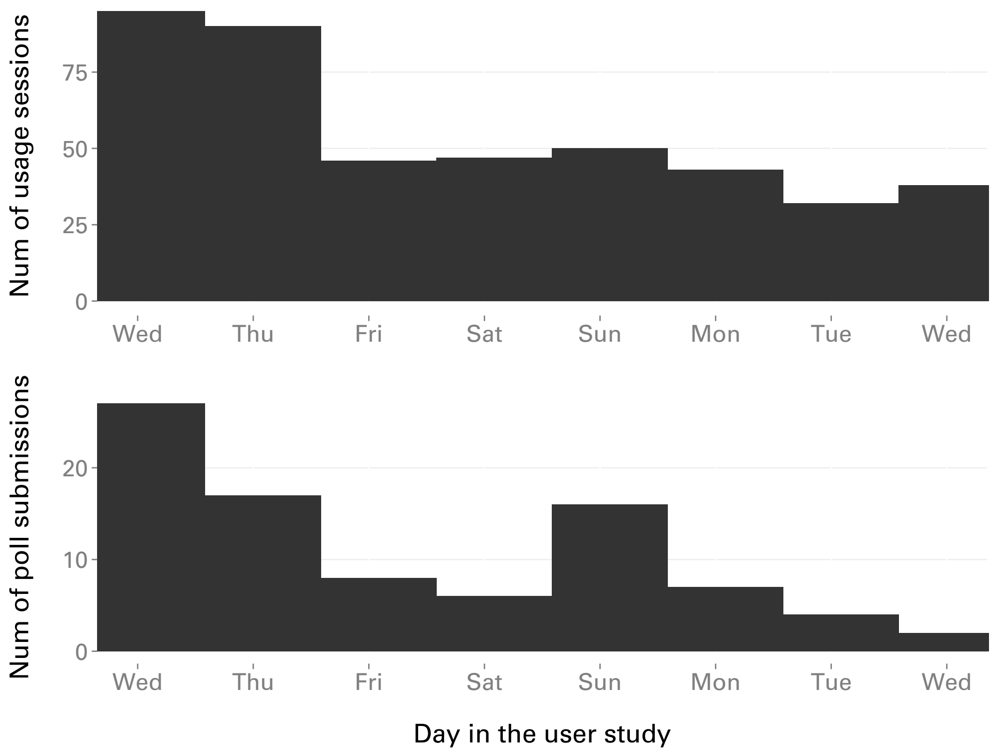


Figure 5.11: Histograms showing user activity (as recorded from their devices) with the FlashPoll app over the course of the study. The top graph shows the number of usage sessions each day; the bottom graph shows how many poll submissions were recorded each day. $N = 40$.

No significant differences were found in terms of which activities participants used their phones for, neither between Android and iPhone users nor by gender, level of education or employment status.

5.2.4.3 How much did participants use the FlashPoll app?

In total, 386 usage sessions[9] were recorded during the seven days of the study ($N = 40$; see Figure 4.6 for a timeline of the study). The median length of a usage session was 20.6 seconds (32.2s for Android users and 13.4s for iPhone users).

As can be seen in Figure 5.11, there was much activity on the first two days, which then dropped on the third day, staying relatively stable until the end of the study. For the recorded number of submitted poll answers, there was a similar story: many were answered on the first two days, and then progressively fewer as the week went on; there was an increase on Sunday, which coincided with the becoming available of a poll covering the whole city of Berlin.

[9] A usage session was considered to be one or more "event logs" from a certain user separated by less than 10 minutes. That is, if more than 10 minutes passed between a user's actions, they were considered to have begun a new usage session.

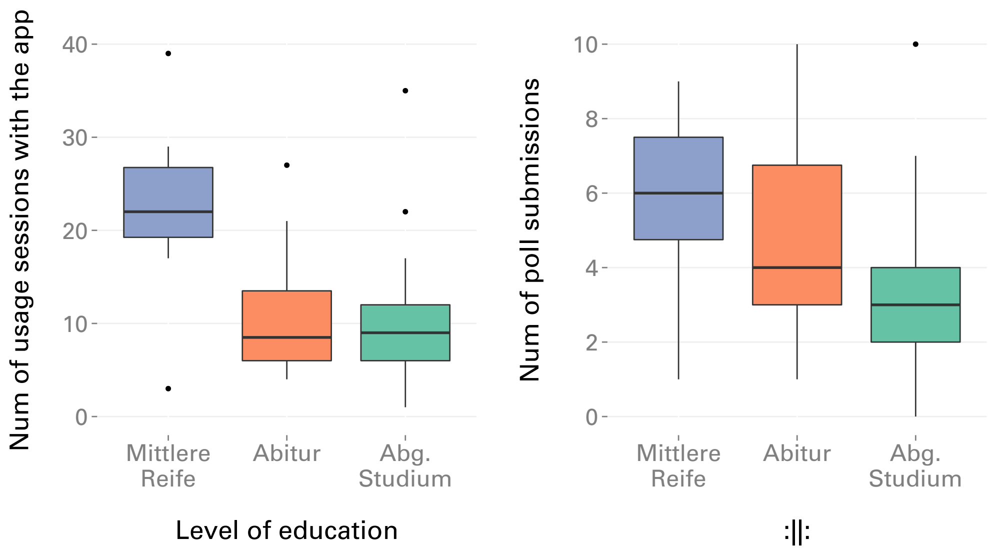


Figure 5.12: Box plots showing differences between participants with different levels of education. The vertical axes show the *number of recorded usage sessions* (left) with the FlashPoll app and the *number of recorded poll answer submissions* (right) in the course of the study. $N = 39$.

No significant differences were found in terms of Android/iPhone users, gender or employment status. However, a one-way ANOVA showed a significant difference ($F(2, 36) = 6.208, p = .005$) in number of usage sessions between participants with different educations. Specifically, a post-hoc test using Mann-Whitney tests with Holm-Bonferroni correction showed that *participants with a lower level of education (Mittlere Reife) had more recorded usage sessions than those with a high level of education (Abgeschlossenes Studium)* ($p = .033, r = .103$; mean number of sessions for the groups were 22.3 and 10.4 respectively). This is illustrated in the box plot in Figure 5.12. No significant difference was found between participants with educations of types Abgeschlossenes Studium and Abitur ($p = 0.062$).

For the recorded number of submitted poll answers, no significant differences were found either with regard to Android/iPhone users, gender, employment status or level of education.

5.2.4.4 When did they use it?

In order to find out when in the day (that is, at which hour) participants used the FlashPoll app, we summed up all the recorded sessions that occurred within each hour period (00:00–01:00, 01:00–02:00, etc.) and plotted them on a histogram (see Figure 5.13). Looking at it, we see usage generally increases throughout the day, albeit with many temporary up- and downturns. The most active periods in the day were in the afternoon, between 13:00 and 15:00, and in the evening (20:00–21:00 and 23:00–24:00), whereas the least active periods were in the night, between 00:00 and 01:00, and during lunch hour (12:00–13:00). Participants nearly used the app as much (or little) in the night (from 01:00 to 05:00) as they did in the late afternoon

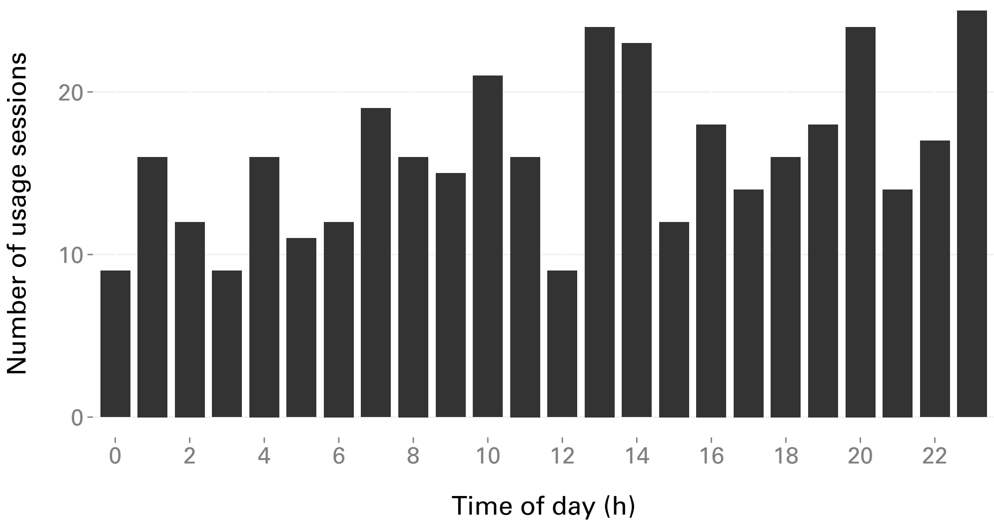


Figure 5.13: Histogram showing the number of usage sessions (as recorded from their devices) with the FlashPoll app for each hour of the day. Participants used the app most frequently in the afternoon (13:00 to 15:00) and in the evening. $N = 40$.

(15:00 to 19:00).

In order to see when participants responded to polls, we look at the answers given in the diary entries. As can be seen in Figure 5.14, participants were most active in responding to polls in the afternoon ($\mu = 3.33, \sigma = 2.01$), followed by in the evening ($\mu = 1.89, \sigma = 1.96$) and in the morning ($\mu = 1.71, \sigma = 1.69$).

No differences, either in when participants used the FlashPoll app or when they responded to polls, were found either between Android and iPhone users or in terms of gender, employment status and level of education.

5.2.4.5 Where or in which context did they use it?

As we can see in Figure 5.15, participants overwhelmingly responded to polls in two locations, or contexts: at home ($\mu = 3.76, \sigma = 2.36$) and while commuting ($\mu = 3.16, \sigma = 2.09$). By contrast, it was rare for participants to respond to polls at at their school/university ($\mu = .34, \sigma = 0.86$) or in their workplace ($\mu = .20, \sigma = 0.66$).

With regard to the number of answers at a certain location, there were no significant differences between groups, either in terms of Android/iPhone users, gender, employment status or level of education. However, two significant correlations were found for iPhone users, which are both illustrated in Figure 5.16.

First, there was a strong, negative correlation ($r(15) = -.536, p = .026$) between iPhone users' *civic engagement* and their tendency to respond to polls *in the home*. That is, for iPhone users, having a stronger civic engagement was correlated with responding to *fewer* polls at home (and less civic engagement with responding to

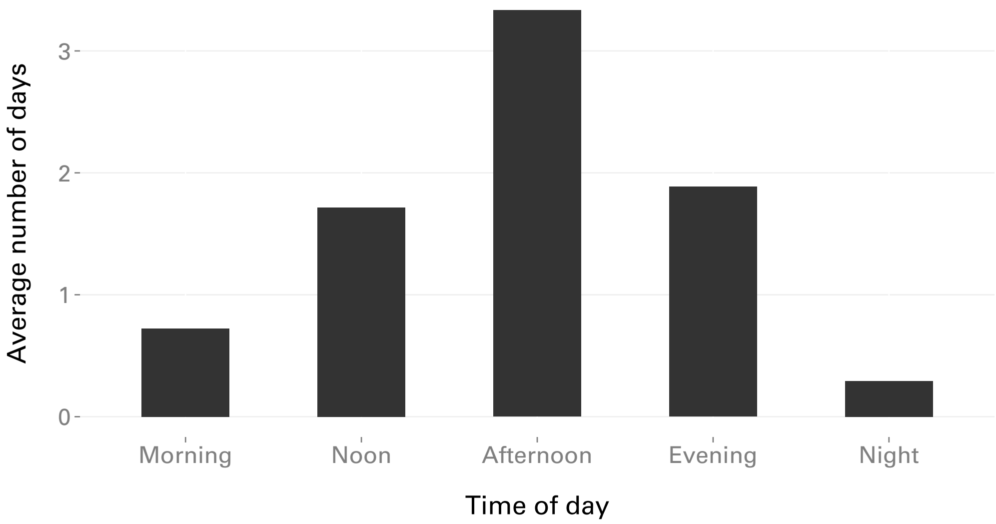


Figure 5.14: Graph showing when in the day participants responded to polls, as reported in their diary entries. The vertical axis shows on the number of days, on average, on which a participant reported having responded to a poll at a certain time. $N = 39$.

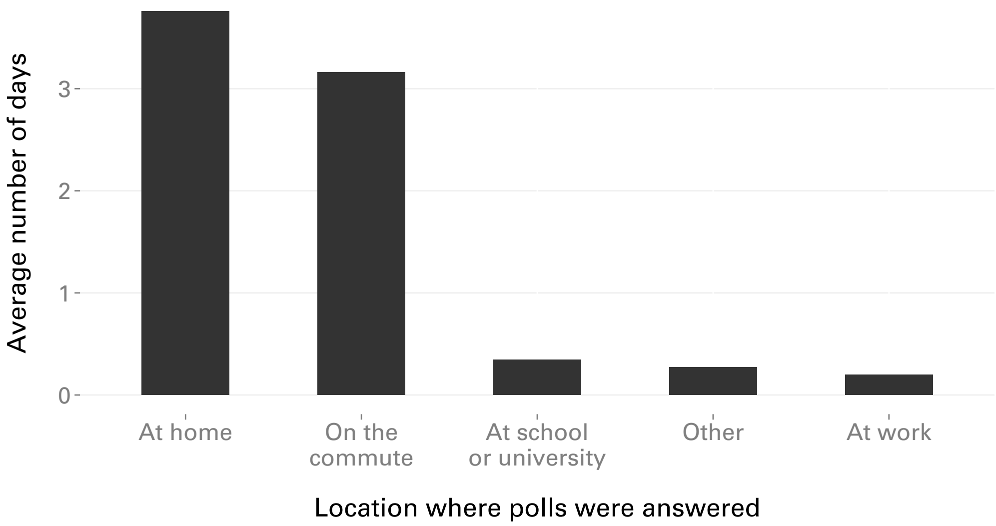


Figure 5.15: Graph showing where participants responded to polls, as reported in their diary entries. The vertical axis shows on the number of days, on average, on which a participant reported having responded to a poll at a certain location. $N = 39$.

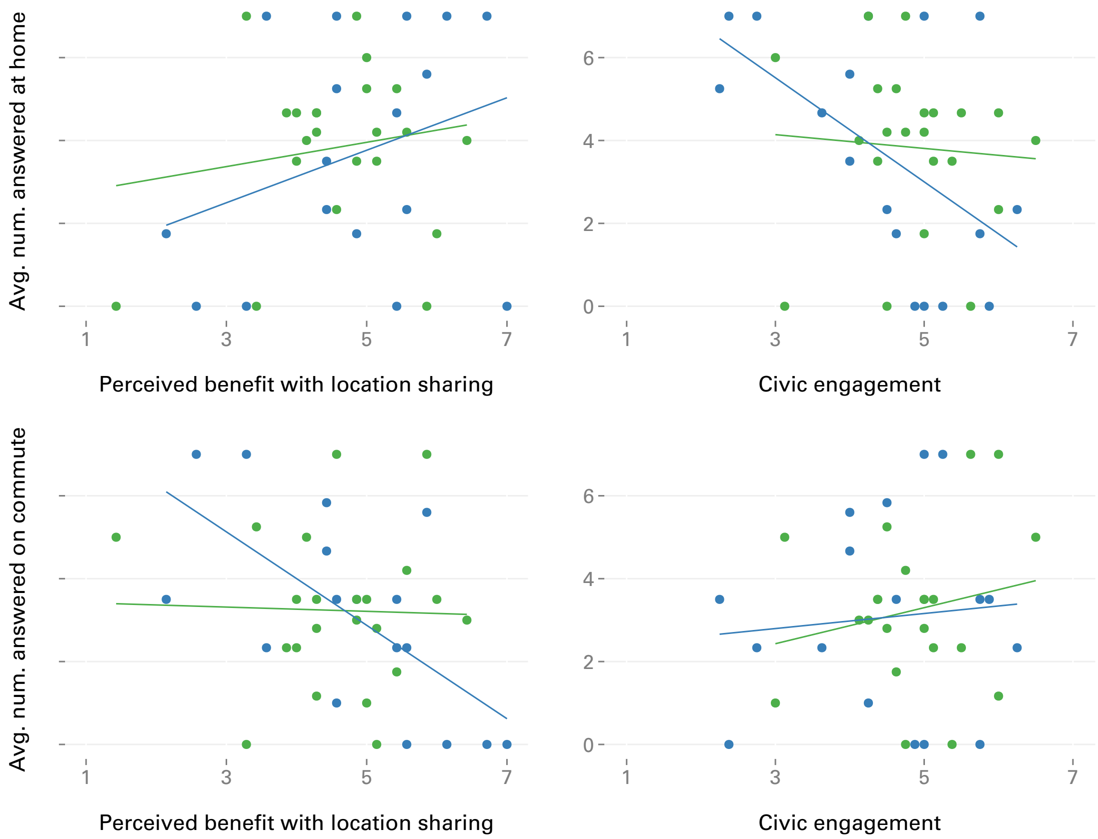


Figure 5.16: Scatter plots illustrating relationships between how much many polls participants responded to in certain locations (vertical axes) and various psychometric measurements (horizontal axes). Each dot is a participant; Android users are green and iPhone users blue. Owners of iPhones responded to more polls at home if their civic engagement was lower (top right), and *fewer* polls while commuting if they saw *more* benefits in sharing their location (bottom left). $N_a = 21, N_i = 15$.

Motivation	μ	σ
Responding to a poll		
“I was bored and responding gave me something to do”	2.27	2.87
“I wanted to try out the app”	2.26	2.31
“The poll’s topic interested me”	2.24	2.54
“I felt obliged or expected to”	2.22	2.77
“I wanted to share my knowledge or opinion”	2.18	2.47
“I wanted to help others”	1.44	2.63
Using the app but not responding to any poll		
“There were no polls available”	2.42	3.33
Not using the app		
“There were no polls available”	2.46	2.87
“I didn’t have time”	2.32	2.65

Table 5.5: Table showing the results for participants’ motivations in usage and non-usage of the app. Includes only those motivations with a mean of 1 or more. See Figures 5.17, 5.18 and 5.19 for visual representations. $N = 39$.

more polls at home). For Android users, there was no such correlation ($r(20) = -.070, p = .755$).

Second, there was a strong, negative correlation ($r(15) = -.640, p = .006$) between iPhone users’ *perceived benefits with sharing location information* and their tendency to respond to polls *on the commute*. In other words, for iPhone users, seeing more benefits in sharing one’s location was correlated with responding to *fewer* polls while commuting, and seeing less such benefits with responding to more polls while commuting. There was no such correlation for Android users ($r(20) = -.029, p = .897$).

5.2.4.6 What were their reasons for using it or not using it?

In order to find out more about participants’ motivations for using or not using the app, we shall consider three things: first, their reasons for using the app to respond to polls; second, their reasons for using the app but *not* responding to any polls; and third, their reasons for not using the app at all.

Looking first at participants’ reasons for using the app to respond to polls, we see (Figure 5.17) a lot of variety, with participants picking some five or six reasons at about the same rate. These reasons (in addition to the other results) can be seen in Table 5.5. It is worth noting at this point that one of the reasons – “I felt obliged or expected to” – does not occur naturally in the using of an app, but rather signifies an artificial effect arising in research. The other common reasons for using the app can be roughly divided between those signifying enthusiasm for the app’s functionality (“The poll’s topic interested me” and “I wanted to share my knowledge or opinion”),

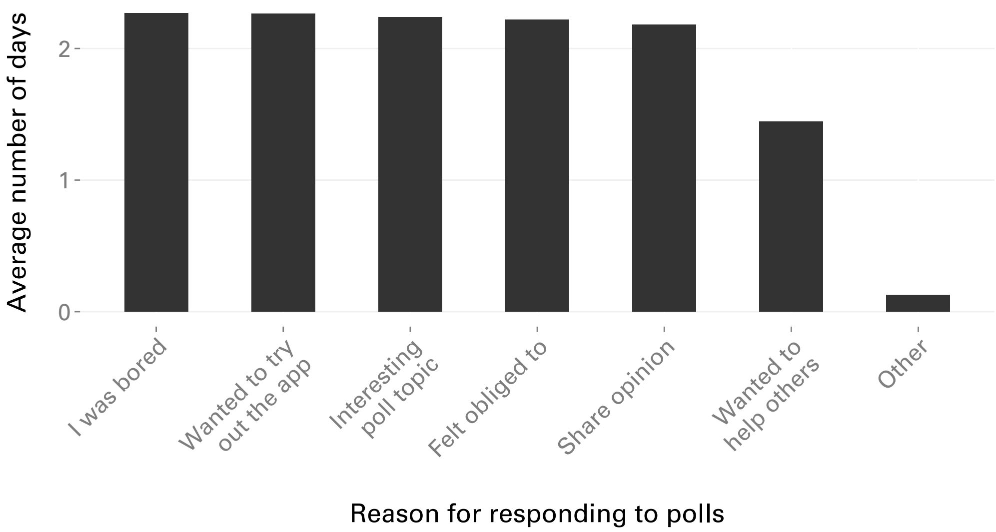


Figure 5.17: Graph showing participants' stated reasons for responding to polls, as reported in their diary entries. The vertical axis shows on the number of days, on average, on which a participant reported having responded to a poll for a certain reason. $N = 39$.

and those representing a sort of disinterested curiosity ("I was bored and responding gave me something to do" and "I wanted to try out the app").

Reasons for *not* responding to polls, on the other hand, were fairly uniform – the only notable reason was, quite simply, that there were no polls available to answer ($\mu = 2.42, \sigma = 3.33$; see Figure 5.18). Note that this concerns those days wherein participant used the FlashPoll app, but without submitting any answers to any of the polls.

Finally, we look at participants' reasons for not using the app at all. Again, as can be seen in Table 5.5, the most commonly recorded reason was, again, that there were no polls available to answer for the user ($\mu = 2.46, \sigma = 2.87$; see Figure 5.19). The second most common reason was that the participant "didn't have time" to use the app ($\mu = 2.32, \sigma = 2.65$). Notably, the app's pragmatic qualities (e.g. its ease of use) seem not to have been an important reason for not using the app; privacy-related reasons, too, were rare.

Here, we find two notable, significant between-group differences. First, as can be seen in Figure 5.20, there was, between users of Android and iPhone mobiles, a significant difference (Mann-Whitney $U = 154, N_a = 22, N_i = 17, p = .046$ two-tailed) among participants who didn't use the app because they "didn't want it to track [their] location". Mean numbers of diary entries in which this option was selected were 0 for Android users and 0.185 for iPhone users. That is to say, iPhone users were more likely to report that they did not use the FlashPoll app because they did not want it to track their location. It should be noted, however, that this reason was overall pointed out very rarely.

Second, as can be seen in Figure 5.21, there was a significant difference between

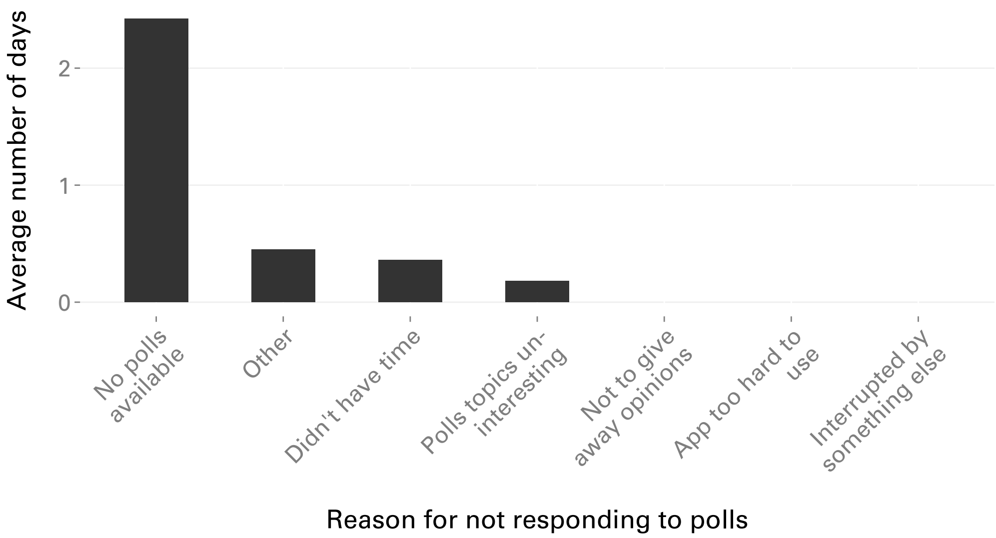


Figure 5.18: Graph showing participants' stated reasons for *not* responding to polls, as reported in their diary entries, *even when they used the app.* The vertical axis shows on the number of days, on average, on which a participant reported not having responded to any polls for a certain reason. $N = 39$.

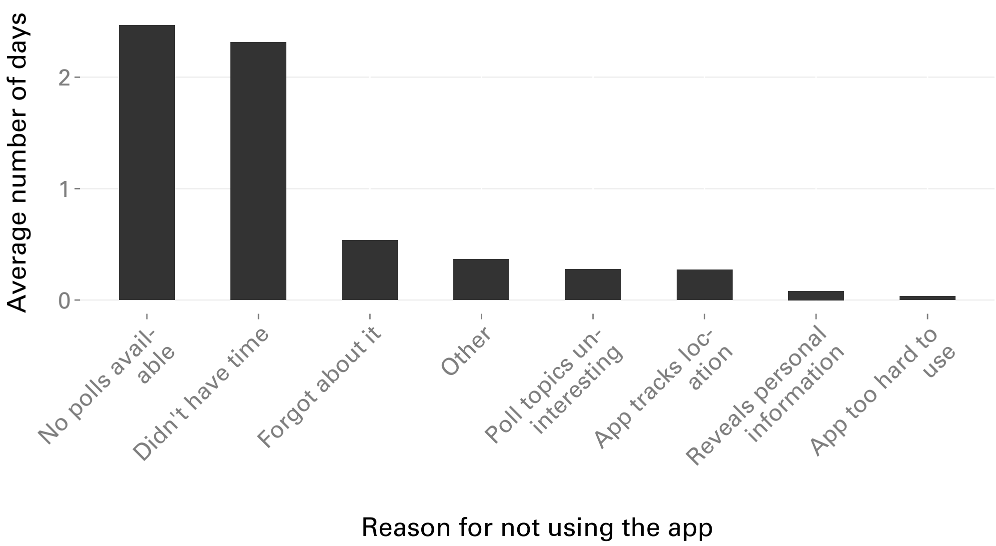


Figure 5.19: Graph showing participants' stated reasons for *not using the app*, as reported in their diary entries. The vertical axis shows on the number of days, on average, on which a participant reported not having used the app for a certain reason. $N = 39$.

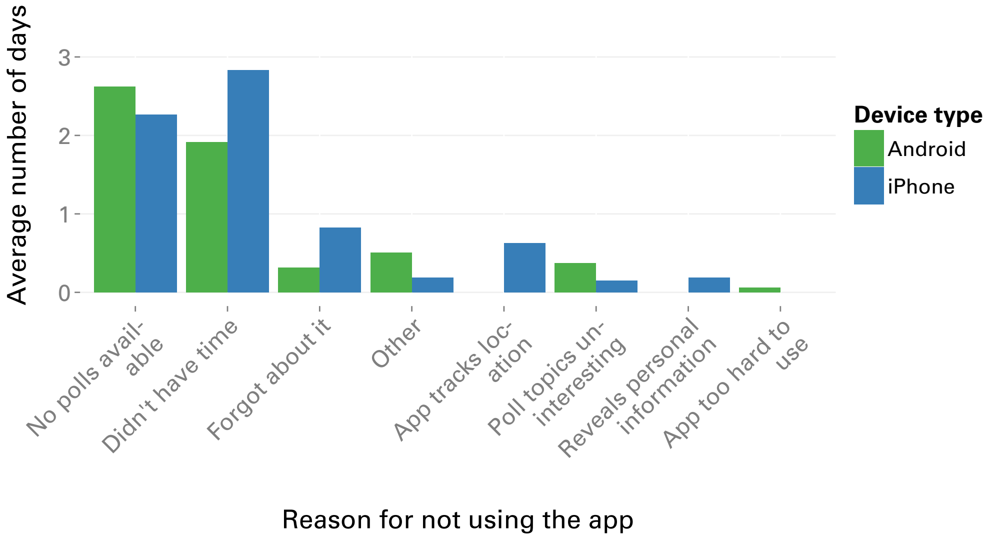


Figure 5.20: Graph showing Android and iPhone users' stated reasons for *not using the app*, as reported in their diary entries. $N_a = 22, N_i = 17$.

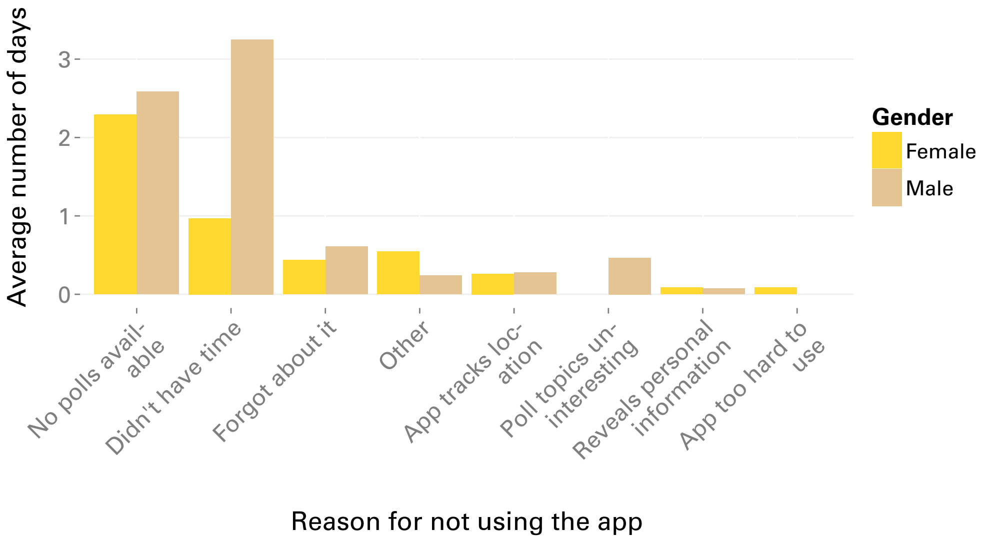


Figure 5.21: Graph showing male and female users' stated reasons for *not using the app*, as reported in their diary entries. $N_f = 16, N_m = 23$.

male and female participants (Mann-Whitney $U = 102, N_m = 23, N_f = 16, p = .014$ two-tailed) among those participants who chose not to use the app because they "didn't have time" to do it. Mean numbers of diary entries in which this option was selected were 3.251 for male and 0.970 for female participants. In other words, male participants were far more likely to report that they did not use the app because they did not have the time available to do so.

5.3 Design guidelines for a mobile polling application

In our introduction (see Section 1.4), we posed the following question: "How to best design the interaction for a mobile polling application for iOS, taking into account the characteristics of the iOS platforms and its users?"

From my experiences with making the FlashPoll app, and using the results of the user study, I have constructed a set of guidelines for the design of a future, similar iOS app. They are divided here into two groups, the first being about interaction and graphics, and the second about user-related considerations.

5.3.1 Interaction

Think of the app as *productory* rather than as for consuming. Unlike, for instance, Apple's Photos or Weather apps – both of which are examples of apps made for consuming content – a productory app exists mainly to create or produce something (in the case of FlashPoll, a response to a poll). True, the FlashPoll app also allows its users to look at results for certain polls, but its main function is still the responding to polls.

This has certain implications for design. Consider the iOS Human Interface Guidelines: "Start by considering the tasks in your app: How often do users perform them and under what circumstances?" (Apple 2015b). The main task, the main course of action in this sort of app is *the answering of polls*.

Have the app guide the user through the main course of action. This main course of action can be thought of as a path through the app's content. It should be promoted (in a subtle, gentle way) throughout the app, from the user's opening of it until his or her having completed the action (submitted a response to a poll). Again, consider the iOS Guidelines: "Users should always know where they are in your app and how to get to their next destination" (Apple 2015b).

The FlashPoll app does this in several ways. In the list of polls that meets the user upon opening the app, active polls (i.e. those that can be answered) are given more prominence than past polls (those that can no longer be answered). For each active poll, the time remaining until it closes is shown, in order to encourage the user to give a response while still possible. Having selected a poll, the user is shown its

Description section, not the Poll section which is where one responds to the poll's questions. In order to guide the user to the questions, the Description section has a "Start" button, which performs that function.

Finally, throughout the process of responding, the app uses animation to keep a "forward momentum": opening a poll brings the poll view in "from the right"; navigating from the Description section to the Poll section also initiates a movement from the left to the right (see Section 5.1.2); and, finally, navigation through the poll's questions is done from left to right, so that the final question is also the rightmost one. The whole course of action thus has a spatial dimension to it, conceptualised as a movement from left to right.

A single bad interaction can frustrate the whole experience of responding. When comparing the Android and iOS versions of the FlashPoll app, I found that users took less time to respond to a poll with the Android version (see Section 5.2.2). This difference was almost *entirely* caused by a single interaction: the drag-and-drop rearranging of options for ranking questions.

The deficiency was noted by several users, who wrote that it "[was] not intuitive enough" and that it lacked effective hints. The interaction in the Android version (which was not based on a drag-and-drop interaction), however, although it allowed for quicker responding than the iOS interaction, was not perfect either, with participants deeming it unclear and "cumbersome".

Conclusion? A polling app is based on each question being easily answerable. Focus and design time, therefore, needs to be allocated also on these "bottlenecks", so that the interaction flows smoothly throughout.

Aim for a sober, serious appearance. Being an app which not only requires the user's location, but which is also dependent on the user's sharing his or her opinions and experiences, it was important that the FlashPoll app inspired trust in its users. Five of the ten participants in the user interface evaluation commented positively on the iOS app's appearance (three did so for the Android app; see Section 5.2.2), writing that it "looked nice" and commending its "colourful presentation". Moreover, eight of the ten participants mentioned appearance either as a positive or a negative, indicating its importance in shaping the overall impression of the apps.

It is not clear whether such an appearance makes users more inclined to respond to polls, but, as the iOS Guidelines states, "an app that helps people perform a serious task can put the focus on the task by keeping decorative elements subtle and unobtrusive", sending "a clear, unified message about its purpose and its identity that helps people trust it" (Apple 2015b).

Make the app remember. Because there is no guarantee that a user will sit down and neatly answer all of a poll's questions in one go (to take an example, many participants in the user study responded to polls while commuting), it is necessary for an app to save the app's state between sessions. This entails remembering their chosen options, their partially written answers, which view they were seeing at the time, etc., and it should not require any action on the user's part.

Keeping the user experience consistent in this way follows directly from the iOS

Guidelines, which state: “Give people confidence that their work is always preserved unless they explicitly cancel or delete it” (Apple 2015b). Cooper et al. (2007) get to the same thing with the phrase: “Smart products have a memory”. This is especially important in a productive app, where users can take quite some time to put down their answers (especially when writing).

5.3.2 Considerations regarding the user

If the app is location-based, show the user the benefits of this. In the user study, there was a correlation for iPhone users between time spent with their phones and the benefits they saw with sharing location information (see Section 5.2.4.1). Moreover, iPhone users were significantly more likely than Android users (although still not very likely) to *not* use the FlashPoll app because it tracks their location (see Section 5.2.4.6).

The implication of this is twofold. One, if an app uses location information, it should ensure that the user knows how he or she benefits from this usage. Two – because there did not seem to be a correlation between perceived disadvantages of sharing location information – this should focus on the *benefits* of location sharing rather than on addressing any perceived risks with it.

If a user sees benefits with sharing location information, he or she may be less inclined to use the app while commuting. Paradoxically, there was in the user study a negative correlation between the benefits an iPhone user saw with sharing location information and the likeliness of that user answering polls on the commute (as opposed to, say, doing it at home) (see Section 5.2.4.5). There was a very slight, positive correlation for answering polls *at home*, but this was not statistically significant.

What that means is that a “side effect” of the user’s seeing more benefits in sharing his or her location is that he or she will be less likely to use the app while commuting, and somewhat more likely to do it at home.

You aren’t necessarily looking to attract users with strong civic engagement. It would be natural to assume that a person with strong civic engagement would be more inclined to respond to polls related to the local area, but this assumption may be a false one.

In fact, nothing in the user study pointed towards users with strong civic engagement using the FlashPoll app more. In the user study, iPhone users with strong civic engagement were less likely to respond to polls at home (see Section 5.2.4.5), and Android users with strong civic engagement spent less time with their mobiles overall (see Section 5.2.4.1).

Moreover, looking at their reasons for using the app, the participants were more likely to select self-focussed reasons (e.g. “The poll’s topic interested me”) than altruistic ones (e.g. “I wanted to help others”; see Section 5.2.4.6). This would suggest that, in order to increase users’ activity, it’s more effective to appeal to their self-interest rather than to ask them to be “charitable”.

6

Discussion

This chapter will discuss the results and the process used in achieving them. It will look at the app in its final state and the results of the user study. It will also bring up some ethical issues related to the thesis and suggest possible future work on the subject.

However, before going into those matters, I will write a word about the "contradiction" inherent in the subject matter.

6.1 The thesis' bipartite nature

This thesis is a thesis about two, quite separate, things: on the one hand, it is about differences between iOS and Android users, and on the other, about the design of a mobile polling application for iOS.

Now, certainly these two topics are related and feed into each other, so to say, but even so each of the two could have been studied quite well separately, without being "mixed up" with the other. That it came about to be so was largely for practical reasons: an iOS app of that sort needed to be made, and thus the opportunity, with two quite equivalent versions of the same app, was there for a comparative study.

As a result of all this, and as will be readily apparent to anyone who reads this report, it, the report, and indeed most of its chapters individually, has a certain contradictory quality. What effect did this have on the work itself?

The main issue that it caused was its threatening to disrupt the cohesiveness of the thesis (such as it was). Even so, several things glued the two subjects together: the app was used in the user study; the user study covered several topics, such as location-related privacy issues, that were relevant to the app's design; the results and experiences of both were considered together when drawing up the design guidelines; and so on.

One can perhaps conclude that, although it may have been the case that each of the subjects could have been more easily studied on its own, the thesis and its report still have a distinct, consistent and unifying narrative thread.

6.2 The FlashPoll application

In Section 6.2 we saw how the FlashPoll app for iOS ended up. Looking at its navigational structure, it was very similar to its predecessor – the FlashPoll app for Android – albeit with some differences, such as in the division between "Active Polls" and "Expired Polls". In terms of graphics, animations and lower-level interactions, on the other hand, it was very different.

Using the results of the user study (which will be discussed in more detail below) and my experiences in designing, developing and using the app, I can identify a couple of problematic areas with it. As found in the user interface evaluation (see Section 5.2.2), there were issues with ranking questions, especially pertaining to the drag-and-drop interaction which, for mainly technical reasons, did not work as well as it ought to have.

Moreover, as noted in Section 4.2.2, the view showing results for ranking questions was also constricted by preexisting technical limitations. In fact, the main problems with the app were related not to its design per se, but to "invisible" problems, such as problems with finding polls while the app runs in the background.

6.2.1 An ambiguity in the design goal

There was a certain ambiguity in the goal of the design of the FlashPoll app for iOS. The ambiguity was this: should the app be as good as possible? or should it be as equivalent to the Android app as possible? Assuming for a moment that the Android version of the FlashPoll app isn't perfect in every respect, these two goals cannot be simultaneously achieved. The FlashPoll project would prefer the former, whereas for the purposes of the user study, the latter would be more desirable.

In the end, the app was designed with the former goal in mind, aiming to be as easy, convenient and pleasant to use as possible. Even so, owing to its having been largely based on the Android app, the two turned out to be quite similar, notwithstanding, of course, platform-specific conventions and various other minor differences. They were especially similar in terms of their overall navigational hierarchies and information structures, and less so in terms of visual appearance and transitions, which was also confirmed by the user interface evaluation (more of which below).

It is impossible to say precisely to which extent this ambiguity has influenced the differences (or lack of differences) found between iPhone and Android users in the study, but it's obvious that the design, development and subsequent study would only have benefited from having had this ambiguity in mind from the very beginning, even if it could not be entirely resolved.

6.3 The user study

6.3.1 Validity and reliability

Validity, as Coolican (2014) writes, concerns "all the likely sources of error in experiments that might cause us to assume, wrongly, that we have demonstrated an effect when none exists, or to assume that an effect didn't occur when, in fact, it did". In other words, we want to know whether the study measures what it intends to measure. In doing so, we shall consider the study's sample, participation rate and issues that cropped up during the course of it. We will also look at the reliability of the psychometric measurements used in it. Finally, we will discuss the extent to which the results can be generalised.

As outlined in Section 5.2.1, the sample had a fairly good demographic spread, with an acceptable gender ratio and participants living in ten of Berlin's twelve boroughs. However, there was a disproportionate amount of students and (perhaps not incidentally) young participants. There was also a considerable number of participants with high levels of formal education.

The participation rate was very good, with 95.7% of diary entries and 100% of end-of-study questionnaires submitted. A small number of the daily estimates of time spent with one's mobile was, as described in Section 5.2.1, very improbable, but these were not considered in the related analyses.

Again as outlined in Section 5.2.1, it was discovered during the study that some of the Android participants had been prompted to update the FlashPoll app to the version on Google Play, thus stopping those participants from seeing the polls made for the study. This may have impacted their usage and responses in several ways, however, given that the number of affected participants was low and that the problem did not last for very long, the impact this caused was likely insignificant.

When it comes to the diary study, it is also important to keep in mind the possibility of there being inaccurate recall – a respondent may not remember precisely how many minutes he or she spent with his or her phone the past day, for instance. Still, the questions did not depend on very detailed reporting; one didn't need to remember the precise hour or location of events, but only an estimate (e.g. "Afternoon" or "At home").

Finally, there is the fact that the study was conducted in Berlin, with German-speaking participants. Berlin is not representative of the world in general, and has its own, peculiar currents, tendencies and discourses. For instance, findings about participants' privacy concerns and other such attitudes may be particular to that city.

Moreover, given that a great deal of material for the study (including questionnaire text) was originally written in English and had to be translated into German, there is the possibility for ambiguities or outright misunderstandings to have occurred. However, no such mistakes were reported either during the pilot or in the course of the study.

6.3.1.1 Reliability of psychometrics

In Section 5.2.1.1, I presented the Cronbach's alphas for the various psychometric measurements used in the study. As mentioned there, they ranged from acceptable (0.7087, although this was after one item was deleted in order) to very good (0.8967 for the "Civic engagement" measure).

Of course, these measurements need to be taken with the usual cautions, including the fact that using a seven-point scale may not make perfect sense for all statements (a "Neither" option can be ambiguous), and the response acquiescence bias (that is, a tendency for the respondent to agree with statements), although as mentioned in Section 4.3.2.1 this was mitigated by using a mixture of positive and negative statements. Finally, there is always the possibility that respondents either are not honest or answer so as to "look good", although the statements in the study do not touch any sensitive or taboo subjects.

6.3.1.2 Generalisability

Given all these caveats, to which extent can the results be generalised? Perhaps, rather than looking to generalise about users, it is more sound (and useful) to look to generalise attitudes and discourses; this, at least, is what Goodman (2008) argues in the context of qualitative research.

Considering, for instance, users' privacy concerns, the results presented here seem to run rather contrary to some of the previous research presented in Section 2.1, with iPhone users seemingly taking privacy issues into account to a greater, not lesser, extent than Android users. This suggests that more research is necessary in order to confirm the existence of these attitudes across contexts.

6.3.2 Findings

In the Results I presented the outcome of the user interface evaluations and the field study. For the latter, I split up the analysis into a section for general demographic findings as well as for each of the research questions set out earlier (see Section 5.2.4); I will adhere to the same structure below.

As discussed in the Process chapter (see Section 4.3.5.3), I used Mann-Whitney U tests, one-way ANOVA tests and Pearson's r measures. Because the Mann-Whitney U test was used, rather than regular t-tests, I only needed to assume that the data was ordinal (not interval), and, moreover, the populations did not need to be normally distributed (although they did need to have the same distribution) (Coolican 2014).

ANOVA and Pearson's r measure, on the other hand, *do* assume interval-level data (Coolican 2014). This was not a problem, however, as all data used was either naturally interval-level or could be assumed to be. For instance, as described in Section 4.3.2.1, Likert scale data could be treated as interval-level since the wording of the choices and the visual appearance of the scales enhanced the notion that the options were equidistant.

6.3.2.1 User interface evaluations

As we see in the AttrakDiff scores presented in Section 5.2.2, participants rated the iOS version of the FlashPoll app higher than the Android version, although the difference was modest. A reason for this may have been that the Android app came first, providing a foundation upon which the iOS app could improve. The difference seems mainly to have concerned the apps' hedonic qualities and their overall attractiveness rather than their pragmatic qualities, which makes sense considering that the apps had similar navigational flows but different appearances.

However, participants did make more errors with the iOS version, almost exclusively when answering ranking questions. This was because the touch area for the drag-and-drop interaction wasn't large enough (for technical reasons), and because visual hints were inadequate, especially for those participants who weren't used to iPhones.

The time measurements are what one might expect given the other results – with the iOS version, it took more time for participants to answer a poll (problematic ranking question), but less time to view the poll's results. There was also a significantly high variance when it came to the time taken to complete the first task (answering a poll) with the iOS app, which likely reflected the divide between participants who owned Android mobiles and those who were used to iPhones.

6.3.2.2 Demographic findings

In Section 5.2.3, we saw both that female participants showed higher levels of *conscientiousness* than did male participants and that participants with an Abitur education had higher levels of *emotional stability* than did those with a lower level of education, Mittlere Reife.

These findings, it must be said, are not of much interest in this study. What is more interesting is that no differences were found between Android and iPhone users. However, this should not be taken to mean that these differences don't *exist* – in order to find that out, more research is required, with larger samples.

6.3.2.3 How much did participants use their mobiles?

Asking how much time participants spent with their phones each day (see Section 5.2.4.1), we found the answer to be 2 hours and 18 minutes on average. The period of most intensive usage seems to have been reported on Saturday, so roughly lasting from Friday evening until Saturday evening, and that of least usage on Sunday, roughly between Saturday evening and Sunday evening. Speculation: this latter result may be because Saturday evening is traditionally an evening of activity (and Sunday morning a time to sleep in).

Two variables were found to be correlated with time spent with one's phone. For Android users, there was a negative correlation between *civic engagement* and *time spent with their phones* – it is difficult to say why this is, but one implication of it is that for community-centred activities (such as the sort of polling done with the FlashPoll app) mobile apps may not, in fact, be a perfect fit.

For iPhone users, there was a positive correlation between *perceived benefits with location sharing* and *time spent with their phones*. This correlation makes perfect sense – a person who sees more benefits with one aspect of mobile technology is more inclined to use mobile devices. What is interesting is that participants' perception of *risks* involved in sharing one's location did *not* affect their usage significantly. This is especially interesting in light of the negativity bias, one aspect of which is that negative things (e.g. risks) have a greater effect on a person than positive things (e.g. benefits). A possible implication of this (as discussed also in Section 5.3.2) is that, if one wishes to convince a user to share their location data, it may be more important to promote the benefits of this sharing, rather than to play down the risks involved.

6.3.2.4 What did they use it for?

In Section 5.2.4.2, we saw that participants used their phone for *communication* nearly every day – not a surprising result, to say the least. In general, the most common activities – communication, entertainment (e.g. browsing the web, watching videos) and trip planning (e.g. finding public transport connections) – are habitual, everyday activities, unlike the least common ones – games, hobby activities (e.g. photography, recording music) and shopping – which perhaps are done only on certain occasions or by a markedly smaller segment of the population.

Shopping was by far the rarest activity of those recorded, owing perhaps to a continuing resistance towards carrying out sensitive transactions with mobile devices.

6.3.2.5 How much did participants use the FlashPoll app?

In looking at how much participants used the FlashPoll app during the study (see Section 5.2.4.3), we saw that they were very active (both in terms of number of usage sessions and number of poll response submissions) during the first two days. After a sharp decline between Thursday and Friday, usage was stable for the remainder of the week. A brief bump in poll submissions recorded on Sunday did not seem to come with any change in overall usage – that is, after the initial peaks, usage remained constant despite a varying number of poll submissions.

Here it must be pointed out that the unit of a "usage session" is quite arbitrary, because data was only recorded as discrete events – should two events separated by 10 minutes of inactivity be considered as part of the same session or as two separate sessions? Moreover, the measurement of it was dependent on participants' access to the Internet during the field study (four participants did not have mobile subscriptions with Internet access).

The sharp, initial decline in poll submissions combined with the brief bump recorded on Sunday (when a city-wide poll became available) both point to a problem with the study which will become more apparent below. The problem is that most participants answered all of the polls that were available to them, meaning that the number of poll responses submitted by a participant was likely more influenced by how much that participant moved around geographically than his or her willingness to respond to polls.

Finally, we saw that participants with a lower level of education (Mittlere Reife) had *more recorded usage sessions* than those with a university-level education (Abgeschlossenes Studium). Again, it is difficult to tell why this is, especially as there were no significant differences in terms of employment status, nor did participants with a Mittlere Reife education use their *phones* significantly more than than did those with an Abgeschlossenes Studium education.

6.3.2.6 When did they use it?

Asking *when* participants used the FlashPoll app, we saw that usage was fairly evenly distributed over the 24 hours of the day, albeit with peaks in the afternoon (13:00–15:00) and with generally more intensive usage during the day than during night (see Section 5.2.4.4). There was a significant trough in usage during lunch (12:00–13:00), possibly because many participants were busy eating at that hour (lunch being a more regular meal than dinner).

We saw also that poll submissions occurred most often during the day, especially in the afternoon. What is perhaps surprising is that, while poll responses were submitted almost exclusively during the day, the app was used nearly as much at night as during the day. This would support the previous speculation, of there is little connection between poll response submissions and overall app usage.

6.3.2.7 Where or in which context did they use it?

In Section 5.2.4.5, we saw that participants almost exclusively responded to polls either at home or on the commute. This makes sense: most people spend much of their time at home, especially free time; and when commuting, one not only has free time, but also moves around, meaning one is more likely to find new polls (for instance by entering into a previously unvisited neighbourhood). Of course, there is also an element of guilt involved in admitting that one uses the app at work or at school, where one is supposed to be busy doing other things.

Moreover, there were two negative correlations for iPhone users – one between their *civic engagement* and their tendency to respond to polls *at home*, and one between *perceived benefits with sharing location information* and their tendency to respond to polls *on the commute*. For the former, keeping in mind that there was no significant correlation for answering polls elsewhere, we can suppose two causes – either participants with strong civic engagement spend less time in their homes, or they are less inclined to respond to polls while they are there. We did see in Section 5.2.4.1

that, especially for Android users, civic engagement was negatively correlated with overall phone usage.

The latter is more curious, especially considering the fact that, as mentioned previously, there was a *positive* correlation between these perceived benefits and the time that iPhone users spent with their phones overall. One would expect someone who sees benefits with sharing his or her location to be more open to using the FlashPoll app while commuting, but, on the contrary, it seems such a user was considerably *less* likely to do so, at least in the case of iPhone users.

It is worth noting here that, although there were no significant correlations between these two attitude measures and overall time spent with one's phone or time spent with the FlashPoll app, participants' tendency to respond to polls at a certain location is still dependent on how much time they spend overall. If, for instance, one of these measures was significantly correlated with time spent with the app, it would not be interesting that it was also correlated with their tendency to respond to polls at a certain location, because it would probably be the same for all locations.

6.3.2.8 What were their reasons for using it or not using it?

The last question regarded the reasons reported by participants for their activity (or lack of it) with the FlashPoll app (see Section 5.2.4.6). Before we continue, it should be noted that participants were motivated to respond to polls at least in part for artificial reasons, that is, reasons caused by the special conditions of the study ("I felt obliged or expected to" and, possibly, "I wanted to try out the app"). This is despite the fact that they were explicitly told at the outset of the study that they were not expected to use the app at all, but should only use it as much as they wished to.

Apart from that, it seems participants were mostly responding to polls for self-interested reasons, such as "I was bored" or "The poll's topic interested me", rather than for altruistic reasons like "I wanted to help others". This would suggest that, if one wishes a mobile polling app to generate many responses, it is more important to appeal to people's self-interest rather than asking them to donate their time for someone else's gain.

Turning to the opposite side of the issue, we asked: what were their reasons for not responding to polls, and for not using the app at all? Again we notice the issue which was raised previously (see Section 6.3.2.5), namely that the most common reason for not responding to polls (or using the app) was that there simply were no new polls available to respond to. This points to a weakness in the study – there were too few polls to get any significant differences: a difference in number of submitted poll responses for a participant was more likely to be caused by his or her travelling around a lot (and thus finding new polls) than any special inclination to respond to polls.

Apart from that, the main reasons for not using the app at all ("I didn't have time" and "I forgot about it") were not directly related to the app itself. In fact, very few participants chose not to use the app because of the app itself or because of privacy reasons.

However, there was a small but significant contingent of iPhone users who did not to use the app because it entailed sharing one's location, at least to a certain extent. Now, this difference was mainly caused by a relatively small number of iPhone users putting this reason down on nearly every day of the field study, but it is still interesting in the light of the previous finding: that, for iPhone users, *perceived benefits with location sharing* was negatively correlated with answering polls *while commuting* – another case where privacy questions affected usage.

Finally, we saw that male participants were significantly more likely than were female participants to select "I didn't have time" as a reason for not having used the app. Did men really have less time available than women? As there were no such differences in business in terms of employment status or level of education, and as there were no significant differences between the genders in terms of actual number of poll submissions, it is possible that this was not about being busy at all, but rather an easy, "default" option to select if no other option fitted.[1]

6.4 Process

The design and development of the app and the carrying out of the user study went off mostly without a hitch, although a few problems did occur. These, along with ways in which the process could be improved, are discussed below.

6.4.1 Design and development

The process of designing and developing the FlashPoll app worked well, although the project would have benefited from more time up front for ideation, concept exploration and more elaborate prototyping. Perhaps the most important parts of development were: 1) that the app was in a functioning state early on, allowing tests to be done, and 2) that it was flexible enough to allow the design, or at least elements of it, to be changed in the course of it.

The design process could, however, have benefited from having been consciously run with an overarching design methodology such as, for instance, user-centered design or Cooper's goal-directed design (Cooper et al. 2007). This would have unified the different methods used throughout and, perhaps, allowed for a better analysis of the users' needs or goals before entering the development phase.

[1] In fact, one analysis of the so-called "leisure gap" in America seem to indicate that men, not women, have more leisure time (Wang 2013). But this, of course, is not directly relevant to the subject at hand.

6.4.2 User study

The logistics involved in the study were handled well: it did not miss any of the necessary materials, and all of the participants were on time for their scheduled starting sessions. The location suited the purpose, with minimal aural and visual distractions.

Before it began, 55 polls of varying subjects and questions were created for the field study (see Section 4.3.5.2). However, as I noted when discussing the results that came from it, in hindsight, the study would have benefited from having more of these polls available throughout the city, allowing for more variation in the number of polls participants responded to. There is also the chance that the polls' contents and subject affected response patterns; however, it is unlikely that this effect was significant – for instance, very few participants chose not to respond to polls or use the app because "The available polls didn't interest [them]" (see Section 5.2.4.6). Of course, participants' view of the app may also have been affected by the contents of the polls; if, for instance, there were any badly phrased or nonsensical questions, that may have given them a negative view of it.

As mentioned in the discussion about the sample (see Section 6.3.1), it skewed young and had a disproportionate number of student. This could have been improved by widening the channels of recruitment, going for instance to neighbourhoods where fewer students live.

Finally, a note on the diary entries that participants were asked to submit once a day during the field study. In order to have as few of these entries missed as possible, each diary entry became available at 16 o'clock in the afternoon and remained online until 15:59 the following day. This had the unfortunate side effect of allowing participants to submit their responses at quite varying times. For instance, if person A responds within minutes of the diary entry becoming available and person B responds a minute before it closes, there is a full day between the two responses, even though they are counted as having been made on the same day. With 280 diary entries, this should eventually even out, but even so, of course, it may have affected the results (especially the reported time spent with one's phone). The final issue which may distort the results is the problem of inaccurate recall, although with diary entries being made daily, this should not have been too much of a problem.

6.5 Research ethics

During the field study, data about usage of the FlashPoll app was collected from the participants' mobile devices. All of the participants in the user study signed a consent form before participating, agreeing to this usage. The data collected in the user study was anonymous – it was tied only to an identification number assigned to each participant. It did not include names, personal identification or any other sensitive information. Data was anonymised before being analysed.

The participants in the study were also not told that the polls that they saw in the course of it were "fake", i.e. created for the purpose by the researchers. However, the study ended with a debriefing (see Section 4.3.5.2) that informed the user of the study's polls having been made in this way and why this had been done.

6.6 Future work

Because this thesis concerns a multitude of interrelated subjects and is quite broad in scope, I will suggest several divergent points from which new research may begin. First, of course, there is the design problem itself; not much research has been done on usability in polling/surveying interfaces, especially in a mobile context, and especially when it comes to very specific, but important, interactions (what is the best way to let users order a list of options by some criteria? how should one present the results of open questions?). Indeed, mobile polling/surveying apps use paradigms developed for desktop apps, or even earlier ones coming from pen and pencil questionnaires (checkboxes, radio buttons) – it may well be the case that they don't fully take advantage of later technologies, such as animation and touch gestures like dragging, pinching and multi-finger gestures.

With regard to the user study, there are a few areas which could benefit from further research.

One, there is the question of what, exactly, drives usage of this sort of community- or society-centred apps. Because, contrary to what one might think, it certainly does not seem to be civic engagement, nor were any correlations found between usage and personality traits like extraversion or openness to experience. The measure of civic engagement is, of course, only concerned with participants' attitudes – perhaps a measure of users' civic *behaviour* can shine a light here. There is also a corollary question: what can such an app do to increase usage and responses? To which extent do, for instance, rewards – either monetary or "softer" rewards like being allowed to see results for the question – affect usage?

Second, there is the question of how users' perception of risks and benefits involved in sharing their location data affects their perception and usage of apps, both mobile apps in general as well as those specifically concerned with polling and surveying. We've seen both that iPhone users who saw more benefits with sharing their location would spend more time with their phones (and vice versa), and that those same users tended to respond to fewer polls while commuting (and vice versa). At the same time, perceived risks of sharing location information did not seem to affect users' behaviour. Why is this? How does one best show users the benefits of sharing their location data? How come users who see more benefits with location sharing actually responded to *fewer* polls while traveling around?

Third, it would be interesting to learn more about the context in which participants use these apps. In the study, I found that participants would mostly respond to polls either at home or while commuting. Now, the FlashPoll app is a special case, since it is location-based and alerts users when new polls become available (which as the user enters into a certain geographical area). Where do, in fact, people spend most time with their phones? Do they use their phones differently (e.g. in longer sessions or more attentively) at home as compared to on the commute or at work or school? This is a subject which would perhaps be more closely tied to sociology rather than interaction design.

7
Conclusion

This thesis researched iPhone and Android users in the context of mobile polling applications. Its aim was twofold: to ascertain how iOS and Android users differ in terms of attitude, usage patterns, needs and concerns in this context; and to determine how to best design such an application for iOS. It was carried out as part of the FlashPoll project, a project aiming to bridge the divide between citizens and decision-makers. In it, decision-makers can create polls (surveys) covering certain geographical areas, which polls then become available to citizens in those areas through the FlashPoll tool (originally a web app and an Android app).

Having researched the subject in general and the FlashPoll project specifically, I designed and developed a version of the FlashPoll tool for iOS, evaluating and refining the design as development went on. A user study was conducted using this app, along with Android version of FlashPoll. The study attempted to answer six questions formulated based on the original research problem. Forty participants were recruited – 18 iPhone users and 22 Android users. Ten of those participants took part in an evaluation of the two apps' user interfaces, and all forty then took part in a week-long field study, during which they were asked to submit daily diary entries.

The user study resulted in a variety of findings. The user interface evaluation showed that participants rated the iOS version somewhat higher than the Android version, although issues with the interaction for "ranking" questions caused users to need more time with it for responding, than they needed with the Android version.

The field study found, among other things, that participants' civic engagement was, for Android users, negatively correlated with the overall time that they spent with their phones (see Section 5.2.4.1), and, for iPhone users, negatively correlated with their tendency to respond to polls specifically at home (see Section 5.2.4.5). That is, in general, stronger civic engagement in a user seemed to go hand in hand with less phone usage, and vice versa.

Another finding was related to how participants' perceived risks and benefits of sharing location information interacted with usage (see Sections 5.2.4.1 and 5.2.4.5). The perceived benefits of such sharing seemed to be related to more intensive phone usage (but also a lesser tendency to respond to polls while commuting), especially for iPhone users; for Android users, these perceived benefits did not matter as much. Meanwhile, perceived risks of sharing location information did not appear to be related usage at all.

Finally, the study did not find any connection between participants' usage of the FlashPoll app and their responding to polls (see Sections 5.2.4.3 and 5.2.4.4) – on a day when poll submissions increased markedly, overall app usage remained constant, and app usage in general was quite evenly distributed over the 24 hours of the day, even at night, whereas participants responded to polls almost exclusively in the morning, afternoon or evening.

Using both the experienced gained from designing and developing the iOS app as well as the results of the user study, I proposed a set of guidelines for future designs of mobile polling applications. These guidelines were divided into two parts: one set regarding the design of interaction and one est about the potential user.

Bibliography

Aday, Lu Ann and Llewellyn J Cornelius (2006). *Designing and Conducting Health Surveys: A Comprehensive Guide, Third Edition.* 3rd ed. San Francisco, CA: Jossey-Bass.

Apple (2012). *Apple Announces Changes to Increase Collaboration Across Hardware, Software & Services.* URL: `https://www.apple.com/pr/library/2012/10/29Apple-Announces-Changes-to-Increase-Collaboration-Across-Hardware-Software-Services.html` (visited on 06/15/2015).

– (2015a). *Introducing San Francisco.* URL: `https://developer.apple.com/fonts/` (visited on 09/28/2015).

– (2015b). *iOS Human Interface Guidelines.* URL: `https://developer.apple.com/library/ios/documentation/UserExperience/Conceptual/MobileHIG/` (visited on 09/08/2015).

AYI.com (2014). *Android Daters Are Promiscuous Redheads, iOS Are Rich Alcoholics.* URL: `http://www.ayi.com/dating-blog/android-users-promiscuous-online-daters-compared-to-ios/` (visited on 10/09/2015).

Bangor, Aaron, Philip T. Kortum, and James T. Miller (2008). "An Empirical Evaluation of the System Usability Scale". In: *International Journal of Human-Computer Interaction* 24.6, pp. 574–594.

Benenson, Zinaida, Freya Gassmann, and Lena Reinfelder (2013). "Android and iOS users' differences concerning security and privacy". In: *CHI '13 Extended Abstracts on Human Factors in Computing Systems on - CHI EA '13*, pp. 817–822.

Bernard, H. Russell (2011). *Research Methods in Anthropology: Qualitative and Quantitative Approaches.* 5th ed. Lanham, MD: AltaMira Press.

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

Chin, Erika et al. (2012). "Measuring user confidence in smartphone security and privacy". In: *Proceedings of the Eighth Symposium on Usable Privacy and Security - SOUPS '12* 1, p. 1.

Chiu, Dave et al. (2014). *Material design: Structure and components.* URL: `https://www.google.com/events/io/io14videos/b35faef0-7abf-e311-b297-00155d5066d7` (visited on 06/15/2015).

Cockburn, Alistair (2007). *Agile Software Development: The Cooperative Game.* 2nd ed. Boston, MA: Pearson Education.

Coolican, Hugh (2014). *Research Methods and Statistics in Psychology.* 6th ed. New York, NY: Psychology Press.

Cooper, Alan, Robert Reinmann, and David Cronin (2007). *About Face 3.0: The essentials of interaction design.* Vol. 3. 3.

Couper, Mick P. (2000). “Web surveys: a review of issues and approaches.” In: *Public opinion quarterly* 64.4, pp. 464–494.

Deutskens, Elisabeth et al. (2004). “Response Rate and Response Quality of Internet-Based Surveys: An Experimental Study”. In: *Marketing Letters* 15.1, pp. 21–36.

Dillman, Don A. (2000). *Mail and internet surveys: the tailored design method.* 2nd ed. New York, NY: Wiley.

Doolittle, Amy and Anna C. Faul (2013). “Civic Engagement Scale: A Validation Study”. In: *SAGE Open* 3.3, pp. 1–8.

Dunne, Anthony (2008). *Hertzian Tales: electronic products, aesthetic experience, and critical design.* 1st ed. Cambridge, MA: The MIT Press.

Featherman, Mauricio S. and Paul A. Pavlou (2003). “Predicting e-services adoption: A perceived risk facets perspective”. In: *International Journal of Human Computer Studies* 59.4, pp. 451–474.

Felt, Adrienne Porter, Serge Egelman, and David Wagner (2012). “I’ve got 99 problems, but vibration ain’t one: A survey of smartphone users’ concerns”. In: *Proceedings of the 2nd ACM Workshop on Security and Privacy in Smartphones and Mobile Devices*, pp. 33–44.

FlashPoll (2014). *The FlashPoll-Tool.* URL: `http://www.flashpoll.eu/en/project` (visited on 06/15/2015).

FlashPoll Project Consortium (2015). *FlashPoll – A Manual for Successful Mobile Participation.*

Fogarty, Gerard J. et al. (1999). “Validation of a Questionnaire to Measure Mathematics Confidence , Computer Confidence , and Attitudes to the Use of Technology for Learning Mathematics Research Design and Methodology”. In:

Gartner (2015). *Gartner Says Smartphone Sales Surpassed One Billion Units in 2014.* URL: `http://www.gartner.com/newsroom/id/2996817` (visited on 09/14/2015).

Goodman, Simon (2008). “The Generalizability of Discursive Research”. In: *Qualitative Research in Psychology* 5.4, pp. 265–275.

Google (2015a). *Material Design.* URL: `http://www.google.com/design/spec/material-design/introduction.html` (visited on 06/15/2015).

– (2015b). *Google Opinion Rewards.* URL: `https://play.google.com/store/apps/details?id=com.google.android.apps.paidtasks%7B%5C&%7Dhl=en` (visited on 10/05/2015).

Gosling, Samuel D., Peter J. Rentfrow, and William B. Swann (2003). “A very brief measure of the Big-Five personality domains”. In: *Journal of Research in Personality* 37.6, pp. 504–528.

Hassenzahl, Marc (2006). “Hedonic, Emotional, and Experiential Perspectives on Product Quality”. In: *Encyclopedia of Human Computer Interaction.* IGI Global, pp. 266–272.

Hassenzahl, Marc, Michael Burmester, and Franz Koller (2003). “AttrakDiff : Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität”. In: *Mensch & Computer 2003. Interaktion in Bewegung*, pp. 187–196.

Hassenzahl, Marc, Robert Kekez, et al. (2002). "The importance of a software's pragmatic quality depends on usage modes". In: *Proceedings of the 6th International Conference on Work With Display Units (WWDU 2002)* 3, pp. 275–276.

Hesselgren, Sven (1954). *Arkitekturens uttrycksmedel*. Stockholm: Almqvist & Wiksell.

IDC (2015). *Smartphone OS Market Share, 2015 Q2*. URL: `http://www.idc.com/prodserv/smartphone-os-market-share.jsp` (visited on 09/14/2015).

Johnson, Jeff and Austin Henderson (2002). "Conceptual models: begin by designing what to design". In: *Interactions* 9.1, pp. 25–32.

Keith, Mark J., Jeffry S. Babb, and Paul Benjamin Lowry (2014). "A Longitudinal Study of Information Privacy on Mobile Devices". In: *2014 47th Hawaii International Conference on System Sciences*, pp. 3149–3158. ISSN: 15301605. DOI: `10.1109/HICSS.2014.391`. URL: `http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6758993`.

Kuang, Cliff (2013). *The Design Battle Behind Apple's iOS 7*. URL: `http://www.wired.com/2013/06/ios7%7B%5C_%7Dredesign/` (visited on 06/15/2015).

Kujala, Sari et al. (2011). "UX Curve: A method for evaluating long-term user experience". In: *Interacting with Computers* 23.5, pp. 473–483.

Langsner, Jason, By Anthony D Miyazaki, and Ana Fernandez (2006). "Consumer Perceptions of Privacy and Security Risks for Online Shopping". In: *Management* 35.1, pp. 1–7.

Lazar, Jonathan and Jennifer Preece (1999). "Designing and implementing Web-based surveys". In: *The Journal of Computer Information Systems* 39.4, p. 63.

Likert, Rensis (1932). "A Technique for the Measurement of Attitudes". In: *Archives of Psychology* 140, pp. 1–55.

Lohmann, Kris and Jörg Schäffer (2013). *System Usability Scale (SUS) – An Improved German Translation of the Questionnaire*. URL: `http://minds.coremedia.com/2013/09/18/sus-scale-an-improved-german-translation-questionnaire/` (visited on 09/24/2015).

Malhotra, Naresh K., Sung S. Kim, and James Agarwal (2004). "Internet Users' Information Privacy Concerns (IUIPC): The Construct, the Scal..." In:

Michotte, Albert (1963). *The Perception of Causality*. London: Methuen.

Morton, Anthony (2013). "Measuring Inherent Privacy Concern and Desire for Privacy - A Pilot Survey Study of an Instrument to Measure Dispositional Privacy Concern". In: *2013 International Conference on Social Computing*. IEEE, pp. 468–477. ISBN: 978-0-7695-5137-1.

Nielsen, Jakob (1995a). *10 Heuristics for User Interface Design*. URL: `http://www.nngroup.com/articles/ten-usability-heuristics/` (visited on 09/06/2015).

– (1995b). *Severity Ratings for Usability Problems*. URL: `http://www.nngroup.com/articles/how-to-rate-the-severity-of-usability-problems/` (visited on 06/15/2015).

– (2001a). *Success Rate: The Simplest Usability Metric*. URL: `http://www.nngroup.com/articles/success-rate-the-simplest-usability-metric/` (visited on 09/22/2015).

– (2001b). *Usability Metrics*. URL: `http://www.nngroup.com/articles/usability-metrics/` (visited on 09/22/2015).

Nielsen, Jakob (2003). *Recruiting Test Participants for Usability Studies.* URL: http://www.nngroup.com/articles/recruiting-test-participants-for-usability-studies/ (visited on 09/26/2015).
– (2012). *Thinking Aloud: The #1 Usability Tool.* URL: http://www.nngroup.com/articles/thinking-aloud-the-1-usability-tool/ (visited on 09/28/2015).
Nulty, Duncan D. (2008). "The adequacy of response rates to online and paper surveys: what can be done?" In: *Assessment & Evaluation in Higher Education* 33.3, pp. 301–314.
Oomen, Isabelle and Ronald Leenes (2008). "Privacy Risk Perceptions and Privacy Protection Strategies". In: *Policies and Research in Identity Management* 261, pp. 121–138. ISSN: 15715736.
Osgood, Charles E., George J. Suci, and Percy H. Tannenbaum (1957). *The measurement of meaning.* Urbana, IL: University of Illinois Press.
Rettig, Marc (1994). "Prototyping for tiny fingers". In: *Communications of the ACM* 37.4, pp. 21–27.
Rohrer, Christian (2014). *When to Use Which User-Experience Research Methods.* URL: http://www.nngroup.com/articles/which-ux-research-methods/ (visited on 09/26/2015).
Schmitz, Carsten (2015). *LimeSurvey: An Open Source survey tool.* Hamburg. URL: http://www.limesurvey.org.
Sharp, Helen, Yvonne Rogers, and Jenny Preece (2007). *Interaction Design: Beyond Human - Computer Interaction.* 2nd ed. Chichester: John Wiley & Sons.
Smith, Aaron (2013). *Smartphone ownership–2013 update.* Tech. rep. Washington, D.C.: Pew Research Center.
Smyth, Jolene D. et al. (2006). *Comparing check-all and forced-choice question formats in Web surveys.*
Stephenson, Neal (1999). *In the beginning... was the command line*, pp. 1–60.
Tolstoy, Leo (1967). *Tolstoy's writings on civil disobedience and non-violence.* Bergman Publishers, p. 115.
Tufte, Edward R. (2001). *The Visual Display of Quantitative Information.* 2nd ed. Cheshire, CT: Graphics Press.
Usability.gov (2013). *System Usability Scale (SUS).* URL: http://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html (visited on 09/11/2015).
User Interface Design GmbH (2005). *Evaluation Report for Product "Demo - B".* Tech. rep., pp. 1–7.
Wang, Wendy (2013). *The 'leisure gap' between mothers and fathers.* URL: http://www.pewresearch.org/fact-tank/2013/10/17/the-leisure-gap-between-mothers-and-fathers/ (visited on 11/10/2015).
Ware, Colin (2004). *Information visualization: perception for design.* 2nd ed. San Francisco, CA: Elsevier.
Wingfield, Nick and Nick Bilton (2012). *Apple Shake-Up Could Lead to Design Shift.* URL: http://www.nytimes.com/2012/11/01/technology/apple-shake-up-could-mean-end-to-real-world-images-in-software.html?%7B%5C_%7Dr=0 (visited on 06/15/2015).

Wixon, Dennis and Chauncey Wilson (1997). *The usability engineering framework for product design and evaluation*. Elsevier, pp. 653–688.

Xu, Heng et al. (2012). “Measuring Mobile Users’ Concerns for Information Privacy.” In: *Icis* Ftc 2009, pp. 1–16. URL: http://faculty.ist.psu.edu/xu/papers/Xu%7B%5C_%7Detal%7B%5C_%7DICIS%7B%5C_%7D2012a.pdf.

Glossary

- **Android** – An operating system for mobile devices. Developed by Google.
- **AttrakDiff** – A method used to measure the usability of an interface (see Section 3.2.1.2).
- **Heuristic** – Guidelines or principles based on experience.
- **iOS** – An operating system for mobile devices. Developed by Apple.
- **Likert scale** – a psychometric scale where items are rated by “agreeableness”.
- **OS** – Operating System, here referring to iOS and Android.
- **Platform** – Used synonymously with OS, i.e. referring to iOS and Android.
- **Poll** – an assessment of public opinion or experience by questioning a portion of the relevant population; also the term used within the FlashPoll project, cf. survey.
- **Survey** – a structured questioning of the public opinion/experience of a sample, cf. poll.
- **SUS** – System Usability Scale, a measure of usability (see Section 3.2.1.1).
- **UI** – User Interface.
- **Xcode** – The integrated development environment (IDE) used to make iOS applications.

A

Appendix 1 – Example statements for psychometrics used in the study

Below are some example statements for each of the attitude scales described in the Process chapter (see Section 4.3.2.1). They are included in order to give a sense of what each scale actually measures. Note that the AttrakDiff measure is in fact not a Likert scale, but a semantic differential scale.

- **AttrakDiff** (Hassenzahl, Kekez, et al. 2002).
  - PQ: “manageable – unruly”, “straightforward – cumbersome”.
  - HQ-I: “integrating – alienating”, “presentable – unpresentable”.
  - HQ-S: “novel – ordinary”, “bold – cautious”.
  - ATT: “inviting – rejecting”, “attractive – ugly”.
- **Big-Five personality test** (Gosling et al. 2003).
  - Extraversion: “extraverted, enthusiastic”, “[not] reserved, quiet”.
  - Agreeableness: “sympathetic, warm”, “[not] critical, quarrelsome”.
  - Conscientiousness: “dependable, self-disciplined”, “[not] disorganized, careless”.
  - Emotional Stability: “calm, emotionally stable”, “[not] anxious, easily upset”.
  - Openness to Experience: “open to new experiences, complex”, “[not] conventional, uncreative”.
- **Proficiency with mobiles** (Fogarty et al. 1999): “I enjoy trying new things with a mobile phone”, “I [don’t] find having to use mobile phones frightening”.
- **Civic engagement** (Doolittle and Faul 2013): “I believe that I have a responsibility to help the poor and the hungry”, “I believe that it is important to financially support charitable organizations”.
- **Mobile users’ information privacy concern** (Malhotra et al. 2004; Xu et al. 2012): “I read carefully all asked permissions before installing a mobile application”, “Companies behind mobile applications should not use personal information for any purpose unless it has been authorized by the individuals who provided information”.

- **Normative privacy beliefs** (Unpublished study): "People who are important to me think that I should use location-based applications", "People whom I care about and who care about me think that I should use location-based applications".
- **Desire for privacy** (Morton 2013): "I am comfortable sharing information about myself with other people unless they give me reason not to", "I'm comfortable telling other people, including strangers, personal information about myself".
- **Perceived risks with sharing location information** (Unpublished study): "I am worried that if I use location-based applications, strangers might know too much about my activities", "Using location-based applications poses a threat to my personal safety".
- **Perceived benefits with sharing location information** (Unpublished study): "Using location-based applications simplifies communication", "Using location-based applications makes communication faster".

B

Appendix 2 – Questions for diary entries

What follows is the series of questions that each participant was asked to respond to on each day of the study. The questions were asked in German translation.

- Estimate how much time (in minutes) you spent using your mobile since the last diary entry.
- Since the last diary entry, what have you used your mobile for? Pick all that apply.
  - Communicating (calling, texting, messaging)
  - Entertainment (browsing the web, watching videos)
  - Games
  - Shopping
  - Organizing (managing finances, tasks or calendars)
  - Trip planning (booking flights, finding public transport connections)
  - Hobby activities (photography, recording music)
  - Other (please specify)

The following questions were asked only if the participant had used the FlashPoll app since their last diary entry (simply opening the app was enough to be considered "use"):

- Where were you when you answered a poll? Pick all that apply.
  - At home
  - At work
  - At school/university
  - On the commute
  - Other (please specify)
- What time of day did you answer a poll? Pick all that apply.
  - Morning
  - Noon
  - Afternoon
  - Evening
  - Night

- What were you doing just before, during or just after answering a poll?
- What did you like about the app?
- What did you dislike about the app?
- What difficulties did you encounter while using the app?
- If you responded to a poll, why? Pick all that apply.
  - I felt obliged or expected to
  - The poll's topic interested me
  - I wanted to share my knowledge or opinion
  - I wanted to try out the app
  - I wanted to help others
  - I was bored and responding gave me something to do
  - Other (please specify)
- If you did not respond to any polls, why not? Pick all that apply.
  - I didn't have time
  - I was interrupted by something else
  - The app was too hard to use
  - The available polls didn't interest me
  - There were no polls available
  - I didn't want to give away my opinions
  - Other (please specify)

The following questions were asked only if the participant had *not* used the FlashPoll app since their last diary entry:

- Why did you not use the app since the last diary entry? Pick all that apply.
  - I didn't have time
  - I forgot about it
  - The app was too hard to use
  - The available polls didn't interest or concern me
  - There were no polls available
  - I didn't want the app to track my location
  - I didn't want to reveal personal information (please specify)
  - Other (please specify)
- What would make you use the app in future?

C

Appendix 3 – Example of a poll used in the field study

Below is an example of a poll made for the user study, this one specifically about a library, namely Universitätsbibliothek TU Berlin. The same poll was also adapted for other libraries, six in total. All of the polls in the user study appeared in German translation.

- **Title:** Universitätsbibliothek TU Berlin
- **Short description:** What do you think about the library?
- **Long description:** Please take a couple of minutes to improve our understanding of what visitors think about the library and how they use it.
- **Questions:**
  - *(Single-choice, required)* How often do you go to Universitätsbibliothek TU Berlin?
    - Every day
    - Once a week
    - Seldom
    - Once a month
    - Less than once a month
  - *(Multiple-choice, required)* What are you reasons for visiting the library? Check all that apply.
    - Use the Internet
    - Borrow books
    - Research
    - Work or study
    - Rest or relax
    - Make printouts
    - Other (please specify)
  - *(Single-choice, required)* How satisfied are you with the service provided by our librarians?
    - Very satisfied

    * Somewhat satisfied
    * Neutral
    * Somewhat dissatisfied
    * Very dissatisfied
    * Don't know

- *(Open)* How can the borrowing and returning processes be improved?
- *(Open)* What would you like to see more of in the library?