How Smartphone Users Assess the Value/Risk Trade-off of Apps: An Observational Study

Mariano Ceccato, Alessandro Marchetto, Anna Perini. Angelo Susi
Fondazione Bruno Kessler
Trento, Italy
{ceccato,marchetto,perini,susi}@fbk.eu

***Abstract*—The rapid and worldwide diffusion of applications for smartphones (apps hereafter) has produced a complex ecosystem composed by users, apps, developers and vendors with sometimes contrasting and sometimes matching interests. In the literature, this ecosystem has been investigated from multiple perspectives with different kinds of empirical approaches, however some crucial dimensions are still unexplored.**

In this paper we adopt the perspective of Requirements Engineering. We are interested in collecting empirical observations on users' perception of the risks associated to apps when they decide about which app to select and install on their smartphone. Which apps' requirements do users consider? How do they evaluate them with respect to benefits, security and privacy risks?

How users decide about this is still unclear. We think that relevant variables and underlying dynamics must be identified before we can successfully conduct large-scale controlled experiments, as it is already done in other fields of software engineering.

This paper presents the design of an observational study proposed to explore how users assess features and costs/risks when installing apps. The experimental design is then validated and adopted in a feasibility study with a limited set of participants. Preliminary findings are summarised in a set of observations and then discussed in terms of their potential impacts on the app ecosystem.

I. Introduction

Along with the rapid, worldwide adoption of smartphones, we have observed a tremendous growth in the number and diversity of applications (apps from now on) available on marketplaces such as Android Play, the Apple App Store and the Amazon App Store. Smartphone users select and install apps, based on their own needs and interests, with just a few clicks.

The new emergent ecosystem, defined by app developers, vendors, and users, has attracted the attention of researchers that investigate this phenomenon from different perspectives, often with empirical approaches. In earlier studies, a business perspective was taken to investigate the relationships of social, cultural and psychological aspects with the adoption of mobile technology and services. For instance, some works ([16], [5]) conducted extended surveys on hundreds of users, in particular exploiting non-intrusive data logging [16]. More recent studies adopted the perspective of how to engineer app delivery platforms and apps. For instance, [10] and [7] studied the security and privacy of apps and their relationship with application permission models. Also in this case survey approaches have been used.

Taking a Requirements Engineering (RE) perspective, an earlier work [1] pointed out how this emergent ecosystem is challenging traditional RE paradigm and methods. This phenomenon is currently fostering discussion in the research community, towards defining a research agenda [15], [11]. Market-driven RE approaches [12] provide a baseline, although they tend to take mainly the vendor's point of view.

Focusing instead on the users, their potential role in app evolution is studied in research aiming at enabling explicit feedback formulation by users in situ, such as the iRequire tool [14], and more generally in new trends on exploiting social media as an enabler for collaborative software development [3], leading to the so-called *social software engineering*.

In our opinion, there still exists too limited knowledge about how users evaluate costs and privacy, security risks against benefits of smartphone apps. We believe that a deeper understanding of how users select apps to be installed on their smartphones could bring useful insights on the role of *social* RE in this context, and more generally on which methods and techniques can support the engineering of higher-quality apps. This defines our research long-term objective and motivates the work described in this paper, which focuses on the design of a first empirical investigation of the phenomenon.

In fact, while survey-based approaches (e.g., Bouwman et al. [5]) seem to be effective for characterising users' behaviour and attitude towards smartphone apps (and new technology adoption in general) at the social/cultural group level, to collect empirical evidence about the app selection process, we first need to observe users while they are deciding which app to select and install. Interview-based studies (e.g., Chin et al. [7]) can request participants to revisit past or hypothetical experiences, making them strongly dependent on user memory or perception. Differently, an observational approach allows us to look at the actual behaviour and decisional process, rather than to the extent of how users remember it.

This paper describes the design of an observational study, conceived with the purpose of exploring (i) why smartphone users decide that they need an app, (ii) how they decide which app to install and (iii) why they discard other alternatives, which provide similar functionalities (referred as alternative apps from now on). Moreover, we present the validation of this design by conducting a feasibility study with a limited number of participants. The analysis of the collected data provides promising insights on the app selection process, which could be of interest to all the stakeholders of the app ecosystem. For instance, we observed a potential problem related to how users perceive privacy and security risks. This may turn into an opportunity for developers and vendors to develop more flexible permission-based security models, making them user-profile oriented and context dependent. Moreover, the result of this study could be of interest for requirements engineers, when identifying the features for different types of *persona* to be used in a user-centered approach to requirements elicitation [8].

The rests of the paper is organised as follows. Section II describes the design of the study in terms of the research questions that drive our study and the procedures for collecting and analysing empirical data. Section III presents the validation of the proposed experimental design, by executing a feasibility study with a few participants. Early findings are interpreted and commented in Section IV, with an initial sketch of their potential impact, and a discussion of threat to validity. Related work is presented in Section V. Section VI contains concluding remarks and highlights future directions.

II. Study Design

In this study, we intend to observe smartphone users while they face the process of installing an app. We are interested in understanding why they decide that they need an app, how they decide which one to install and why they discard other, potentially alternative apps. Our aim is to uncover the underlying users' evaluations of the trade-off between benefits and risks (particularly privacy and security) of the apps they select.

To achieve this objective, we formulated a set of research questions and identified an appropriate empirical study approach among those available, namely survey, case study or controlled experiment [17], to address those questions. We took inspiration from a recent study [13] that combined observation and interview to investigate how developers perform software comprehension during software maintenance tasks. Observation was used to identify what developers do, while working in their real environment. Interviews helped the authors to understand the motivation behind developers actions. The study explored what strategies developers follow, what type of information they exploit or miss, and what tool they prefer to use. In that case, an observational study was considered appropriate, because the objective of the study called for realism, beside replicability. Results of the study included the confirmation of observations from other studies (conducted with different methods), but also new findings that pointed out an unexpected wide gap between research and practice in software comprehension methods and techniques.

The same requirement of realism and replicability is indeed a key point in our case. However, our study brings additional complexity due to the fact that the user decision-making process is still a poorly investigated phenomenon, so there are no widely accepted metrics to be used with quantitative experimental techniques.

The proposed study consists of two main parts. In the first part, we observe participants when they install apps. In the second part of the study, we ask participants some questions to gain a deeper understanding of what they did and why they did so. We use a structured approach to extract observations from the transcripts of the sessions.

A. Research Question

This study has an *exploratory* purpose [2], and we are interested in investigating the following research questions:

- **RQ**$_1$ How do users select the app to install among those available?
- **RQ**$_2$ Does the selection process change among different users?

Accordingly, we elaborated the study design that is described here below.

B. Data Collection

To meet as much as possible the requirement of realism, we set up our study in a daily life scenario where a user faces the process of deciding which application to install among available alternatives for her smartphone. We do not prescribe the domain/category of the applications to consider, but we let the participant choose a domain that is important and known to her.

Our research method includes both observation and direct questions, as in [13]. We observe what the participant is doing during the experimental sessions and we accurately take note of her actions and of her comments. During the observation we ask participants to comment on their actions. Moreover, for direct questions, we adopt a structured interview format, i.e. we use a predefined list of questions to be asked to the participant, although we do not stick rigidly to such a list. Questions are mainly a guideline to follow during the interview to be sure to record all the relevant topics, however the participant is allowed to freely move from one topic to the other without interruptions. New questions are asked when the participant stops talking as for instance when an argumentation is completed.

Each experimental session is attended by at least two analysts. One of them, namely the *interviewer*, poses questions and interacts with the participant. The other one, namely the *observer* participates in the interview, but she just observes the session and takes note about what is happening. To provide a reliable recording of the experiment, every session is taped using a digital voice recorder. Each experimental session is structured in the following four parts:

a) ***General introduction:*** Each experimental session starts with a general introduction to make the participant comfortable with the study. We start by clarifying the objective of the study, although without explicitly revealing the research questions. In the introduction, we also inform the participant about the scientific research purpose of the study and we commit to keep data confidential. To make the participant feel at easy, we explain that there are no right/wrong answers, so she will not be judged based on her responses. Eventually, we ask the permission to tape the session.

*b) **Observation**:* In order to observe the process of selecting among candidate apps, we start by asking some questions about the prominent use of the smartphone (e.g, work, leisure, entertainment, ...) and if the participant has any particular hobby or interest, such as photography, sports, movies, comics, and so on. The purpose of this initial part is to identify a domain that is known and interesting to the participant, where the participant can clearly identify requirements that are real and relevant for her.

The session continues by asking the participant to suggest and show us a smartphone app that could be useful in this domain. In case the application is already installed in the phone, we ask if alternative apps are available. At this point we observe what the participant is doing, what source of information she accesses (official app store, the web) to identify a new application. In the case the participant decides to install a new/alternative app, we take note of what considerations she expresses to make this decision. In case no new application is installed during the study, we ask the participant to explain why the application already installed is satisfactory, why it was chosen in the past among alternatives and why the available alternative ones were not considered satisfactory. Either cases provide valuable information about the evaluation/selection process adopted by a participant.

To achieve our data collection objective, we adopt the *think aloud* approach. We ask participants to describe aloud what she is doing, what features she is considering and what considerations she is formulating. We ask new questions if the participant stops talking, to re-initiate the information flow. Questions are in the form *"what are you doing/reading?"* and *"why are you doing this?"*. We always pay great attention not to interrupt or influence the participant evaluation process.

*c) **Interview**:* After the observation, the actual interview starts, to understand the reason for the participant decisions and actions. The questions are specific to the class of apps considered by the participant because we do not aim to ask the participant to formulate general rules or abstract considerations. All the questions are specifically referring to the just concluded observation, because we want to collect data on a real scenario. However, even if the starting point is a concrete case, the participant may formulate general sentences. We record general considerations only if they are still applicable to and supported by what we observed.

The interviews are structured around these topics:

- **Decision process** What are the considerations that you formulated to make the decision to install (or not to install) the app during the *observation* part?
- **Comparison** Did you consider other candidate apps? Why were they considered as potential candidates? What features/qualities did you compare?
- **Context** Have you ever been in a different context (e.g., different phone, time or location) where you would have taken a different decision about this app?
- **Security** Do you trust the app vendor and developer? Why? Did you consider the security/confidentiality implications of installing new software on your smartphone?

The interview starts with questions on the adopted decision process, to understand what the considerations that made the participant make the final decision of installing or not installing a new app are. Then, as there are many almost equivalent apps, we are interested to know how participants compare alternatives. For example, they could trust reviews and comments or they could compare apps based on features and, in this case, we are interested in knowing what features are considered. The subsequent set of questions are devoted to investigating what the influence of the context on the decision is, for example in an emergency situation a user might accept to pay more than usual or to accept confidentiality threat if allowing to rapidly solve an urgent problem. Eventually, we explicitly ask if any security-related consideration was taken into account. We mention computer security only at the end of the interview, to avoid that this topic worry or bias the participant and, thus, influence the rest of the interview.

*d) **Profiling**:* The participants of our study are smartphone users that actively download and install apps on their devices who are selected along a suitable sampling policy (see [4]). Our objective is to observe a phenomenon and to formulate an answer to the research question of Section II-A, letting specific aspects that may recur in user decision making to emerge, thus providing candidate measurable indicators for future quantitative empirical studies.

In the last part of each session, we profile participants, initially with respect to age, work position, seniority and gender. Then, other traits of the participants are collected, such as:

- Exposition of confidential data on social networks; this would let us know how much the participant cares about the confidentiality of personal data.
- Exposition to risk, in particular with respect to money; we ask the participants if they make on-line shopping, how often, and if they protect their transactions (e.g., with passwords, insurances or others).
- Knowledge/understanding of computer security; this is important to understand if the selection is a security aware process or guided by unawareness/fear.
- Smartphone familiarity; eventually we are interested to know the level of familiarity with smartphones, i.e. how many applications are installed, how often new applications are installed and when the participant became a smartphone user.

C. Data Analysis

We adopt a structured strategy to analyse data collected during the experimental sessions. After each session, the interviewer and the observer compare and integrate their notes possible mismatches are discussed and solved. The observations formulated by one of them need to be confirmed by the other in order to be included in the final document.

The integrated notes are then summarised into short sentences. Each short sentence is later tagged with labels that represent the main meanings of the sentence. At this stage, labelled summaries of different sessions (corresponding to different participants) are compared. Labels occurring at least in two summaries are reported as a common *concept*.

Observations are formulated first of all starting from concepts occurring more frequently. More observations are also formulated when similar or totally different answers are given to the same questions.

To increase the reliability of our observations, we put in place some guidelines among those suggested by Creswell [9]:

- *Independent peer observations:* In order to limit the threat due to experimenter bias in the observational study, every session is attended by two authors with two different roles, the interviewer and the observer. Any piece of raw data (what the participant does or says) can be used in the analysis only if both the attenders agree on it.
- *Triangulation:* Our study design involves two data sources, they are the observation of participant actions (what they do) and the interview transcripts (what they say). The triangulation of different data sources limits the threats to the reliability of our observations.
- *Participant checking:* We check our observation with participants. They are asked to read and comment a preliminary draft of the data analysis. Their feedback was considered to validate data analysis (Section III)

III. Design Validation and Findings

The experimental design presented in the previous section has been validated through a feasibility study. For the execution of the study we adopted convenience sampling (see Bhattacherjee [4]) i.e. we asked four colleagues to participate (later referred as participant P1, P2, P3 and P4). We shall note that this study does not involve a treatment group and a control group, and that the objective of the study does not rely on statistical analysis to analyse the effect of a treatment (as we did in other studies [6]).

Each session required one hour for the interview, plus an additional hour for the immediately following debriefing between the *interviewer* and the *observer*. Eventually, half and hour was required by a participant to check our preliminary data analysis.

This section collects four *observations* that we could formulate, based on the collected empirical data. We list the *pieces of evidence* (**PoE**) that support each observation. When relevant, we report and cite what was stated by participants between double-quotes and in italics.

Observation 1 *Before installing an app, users have to face an evaluation/decision process.*

This observation is supported by the following pieces of evidence:

– **PoE 1.1** *Apps were installed to satisfy a contingent need or for curiosity.* There can be multiple reasons for installing an app on a smartphone. In some cases, we observed that they were required to solve a specific need that suddenly arose (finding a restaurant for participant P1, travel paperless for P3) or just for the curiosity of trying, for example because the app was suggested by friends (P2 and P4).

– **PoE 1.2** *A selection strategy was adopted.* Independently from the reason why users installed apps, to decide which one to use, participants adopted a quite elaborate (and personal) selection process. P2 used keywords to search in the official market, then looked at reviews from other users and focused on an app with many stars (i.e., high user-based score). She installed it by typing the password and she immediately started the application. When the applications asked special permissions, P2 did not grant such permissions, so she blocked the application and then removed it.

– **PoE 1.3** *Apps were selected as the extension for smartphones of known and trusted services already used on the desktop computer.* All the participants selected apps (or claimed that they usually select apps) that are the porting for mobile phones of a service/application that they already used. Alternatively, an application can be part of an ecosystem of interoperating apps that they already used. In particular P1 stated *"I start from an app that corresponds to a website that I trust and I'm familiar with"*.

– **PoE 1.4** *Opinions of trusted users were used to select the apps to install.* Novice users especially based their decisions on the advice of friends or trusted expert users, possibly as a step of a more structured process. P2 said that *"GoodPdf was suggested by friends"* but also *"Additionally, I ask to people that I know, such as when a friend suggested a map applications with offline maps and good database of point-of-interests"*. P4 reported that she installed *"Whatsapp by friend suggestion, just before a travel"*.

Observation 2 *Installing an app requires evaluating the trade-off between the offered features and their cost.*

This observation is supported by by these pieces of evidence:

– **PoE 2.1** *When selecting an app, participants assessed where personal data will be stored.* An important property to consider when selecting an app to install is the storage of personal data, because it can be local to the phone or remote in a server controlled by the provider. Remote storage was considered a good feature by P4 to preserve data on device failure/theft, while it was considered as a threat to data confidentiality by P1 and P2. P2 said *"I prefer this app because, even if it has less features, it does not save my contacts to the server"* but also *"I like this book reader, because it uses the cloud and it is clear what is on the cloud and what is on the device"*. In particular, P1 was scared of possible misuse of personal data by the application provider, to profile her and send ad-hoc advertisements without her consent.

– **PoE 2.2** *Available features were considered when deciding which application to install.* Most of the attention of participants was devoted to understanding if an application provided the features that were searched for. The features were context-dependent. For instance for P1 the app should find the restaurant fast: *"I installed the application because the* *feature is useful to find a restaurant when I'm abroad and I need to find a typical restaurant that is not too expensive in a short time, such as 30' "*. P2 said that when she searches for an application *"I start from the feature of interest, for example editing, then I refine the results based on feedbacks"*. P3 said *"Applications are different in terms of features, such as capacity, sharing with other people, versioning and cost. I'm not interested in versioning [...]. I'm interested in sharing and capacity."*

– **PoE 2.3** *Permissions requested by apps were assessed as a cost and weighted with respect to other needs*. Participants were aware that apps require permission to access sensitive information and the consequences of this were considered when deciding whether to complete the installation. P2 preferred not to install a feature reach app because it requested too intrusive permissions (the phone address book would have been copied to the cloud) and P4 did not install privacy threatening apps: *"I do not accept if the app tries to access the phone book or my personal data"*.

Conversely P3 installed privacy sensitive apps, because she knew exactly what data were accessed. In particular P3 said *"Privacy is not a big concern for me, because most of the content is not important. [...] I know that they can read it, but nobody would care about these data."*

Sometimes, participants faced a trade off between permissions, features and costs. P1 was aware of a relevant energy consumption due to GPS position capability, but she had to accept it even if this could drain the battery and prevent her to use the phone when it would have been really needed. Conversely, P3 was more concerned about the cost, he stated: *"I'm also interested in cost, dropbox is more expensive than google-drive above a certain threshold."*

– **PoE 2.4** *Multiple alternatives were sometime installed*. When the participants were not able to find an application that fully satisfied them, multiple applications were installed for solving the same problem. P3 said. *"I decide which one to use depending on the type of file to store. Google-drive converts everything to their format, it lets me edit files online and converts pictures into text. Dropbox is fine for all types of document because it does not care about the type"*. But she was unable to decide for a single app, so she stated, *"I have all of them installed in my phone"*.

Similarly P2 admitted *"I chosen Ibook, because it belongs the Apple ecosystem. I chosen Kindle because I already have it on the PC. Goodpdf was suggested by friends"*.

– **PoE 2.5** *Different app-stores were used*. While some users accessed only official stores (P2 and P3) others used non-official stores. In particular, P1 used the unofficial app store pre-installed on her phone by the carrier because she did not realise that it was not an official store. Additionally, P4 prefers to download apps on the desktop computer and to upload them later on the smartphone connected by cable.

This is quite surprising, as P1 and P4 were classified as *conservative* users.

Observation 3 *Users are aware that installing apps involve some (security) risks.*

This is supported by the following pieces of evidence:

– **PoE 3.1** *Participants were aware of the threats to the confidentiality of their data*. Participants knew that apps may have compromised the confidentiality of their data. However, while users with a more conservative behaviour (P1 and P2) perceived this as a risk, other more brave users (P3) considered this an important sharing feature. In fact P1 read the service agreement carefully, with particular attention to the personal data, to how privacy was managed and to the payment method (Google Wallet). While P3 said *"Reliability and sharing is more important than privacy"* and *"I do not store there sensitive data, because people with access rights can access them"*.

– **PoE 3.2** *Big players were more trusted than small companies*. For installing an app that accesses sensitive data, smartphone users had to trust the app and the app provider. Users tended to trust more well-known and big companies than small new companies, because they thought that a big company would not misuse personal data and run the risk of compromising its reputation, and thus its market share. In fact, P2 said *"When I buy applications, I trust the vendor because of its history. I use its cloud since 5 years and I never knew about problems"*. P4 agreed with this point saying *"I pay attention to the producer. If it is a big company I trust it, while if it is a small one I don't trust and I check better"*.

– **PoE 3.3** *Smartphones were not trusted by participants*. Both conservative and brave users distrusted smartphones, but while the former ones were quite afraid of potential problems and limited their interaction with the device, the latter ones used them quite intensively but controlled what data were inserted, because they could potentially leak. P2 said, *"I don't use banking applications on the phone, because I don't trust them. But on the PC I use them"*. P3 said *"I do not store there sensitive data, because people with access rights can access them."* and *"I do not trust the service agreement, because the government and police can always ask to see the data, as anyone can be a suspect." "The information that I put there is not life killer"*. Eventually, P4 admitted *"No payment with the phone"*.

– **PoE 3.4** *Hidden costs were a source of distrust*. Some users did not trust their mobile phones. Among the reasons why users limited their interaction with mobile phones was the perception that there could be hidden costs not declared by the service provider or not evident in the service agreement. For example, even if P1 intended to carefully read the service agreement, at a certain point she said, *"The document is too long"* and to quickly proceed she accepted it without reaching the end. Then, when installing an app she stated, *"They are claimed to be for free, but I do not trust this claim"*.

Observation 4 *Participants have different profiles.*

Participants were colleagues working in our research institute, with different seniority and different familiarity with smartphones. Based on the answers to the profiling questionnaire, this is the way participants described themselves:

– **P1** the *reachable but conservative*: P1 is a senior researcher with fair knowledge of web technology and computer security. Her main reason for using a smartphone is to stay in contact with, and be reachable by, close family members. As such, she never installed apps on her smartphone and she seldom used the pre-installed ones. She exhibits very conservative behaviour, because she permits a very limited exposure of her personal data on social networks, mainly for work reason (i.e., on LinkedIn) and she makes online purchases only with protection mechanisms (pin-protected debit cards or insurance-covered credit cards).

– **P2** the *conservative reader*: Similarly to P1, P2 is also a senior researcher with fair knowledge of web technologies. She avoid exposing personal information to social networks and rarely makes online purchases, and only if providing protection of sensitive data. However, she defined her knowledge of computer security as *"irrational perception and conservative behaviour"*. Her main use of the smartphone is as reading device.

– **P3** the *well-aware exposed veteran*: P3 is a junior post-doc that defines herself as a *"gadget fan"* and her main use of the smartphone is a *"social hub"* to access all her many social networks, where her personal information is highly exposed. Secondarily, the smartphone is a way to be paper-less and environmentally friendly. She is an aware smartphone user, she knows well the technology behind it, and the computer security problems related to it. She frequently installs and removes many apps and she was among the very first smartphone users. She makes online purchases very often using all payment mechanisms (although those with protection are preferred).

– **P4** the *diffident novice*: P4 is a junior post-doc that just recently became a smartphone user, in fact she uses the smartphone mainly to make phone calls. She has a good knowledge of internet technology and a fair knowledge of computer security. She makes online shopping (always protected) but always from a regular computer, never from a smartphone.

IV. Discussion

A. Interpretations

Based on the observation collected in the previous section, we can formulate the following interpretations.

Discussion about RQ$_1$

– *Users start from what they already know:* Smartphones had an overwhelming but recent diffusion, so users are still sceptical and suspicious when perceiving risks in using apps. We observed that participants started their interaction with apps by installing and using what they already know somehow, for example with apps that are the smartphone versions of known services (Piece of evidence 1.3) or those suggested by friends and relatives (Piece of evidence 1.4). Only when users maturated to a certain degree of familiarity and became more familiar with the smartphone did they start to be autonomous and curious and look for apps to address a contingent need (Piece of evidence 1.1).

– *Users trade off between features and cost to select an app:* Often, many almost equivalent apps are available, so users have to decide which one to install. To make this decision, participants adopt a structured, although personal, strategy (Piece of evidence 1.2). In particular, they consider if the features offered by the apps are satisfying (Piece of evidence 2.2) but also if the requested permissions are an acceptable risk and costs (Piece of evidence 2.3). However, the solution of this trade off is not simple and many potential equilibrium points between the two contrasting tensions can be identified. In case many apps equally balance features and costs, more than one alternative app is installed (Piece of evidence 2.4), potentially from different app stores (Piece of evidence 2.5). In the trade off evaluation, however, big and well-known companies are preferred (Piece of evidence 3.2).

– *Smartphones are not trusted, in particular with respect to data confidentiality:* Participants do not trust their mobile phones (Piece of evidence 3.3), they are scared about the risk that apps could disclose and/or misuses their personal data (Piece of evidence 3.1), so they carefully consider whether to install apps or not. To protect and preserve the confidentiality of their data, users take into account the accessed data and the permissions that an app asks for (Piece of evidence 2.3) speculating on the potential risks for personal data (Piece of evidence 2.1). A second concern for user for distrusting smartphones is missing control of what could happen without their notice, such as the presence hidden costs (Piece of evidence 3.4).

– *The context influence the decision of installing apps:* The balance between features and their cost (Piece of evidence 2.3) can shift. Emergency situations for example might require to accept larger costs/risks than normal situations, as the availability of certain features became critical. For example, different alternative apps are sometime installed (Piece of evidence 2.4), to be potentially used in different contexts.

Discussion about RQ$_2$

– *The selection process changes on different users:* Different participants have different characteristics and they adopt different behaviours (Observation 4). So, we expect different users to have largely different sets of installed applications, in fact they are supposed to install apps to satisfy different personal needs (Piece of evidence 1.1), at least because they have to access different services (Piece of evidence 1.3). Similarly, also the trade off between features required and costs accepted by different users (Piece of evidence 2.3) are expected to be different, because it would reflect the *risk profile* of distinct individuals.

Some of these interpretations confirm what already reported by a related study [7], such as the role of advice by friends in the selection on what app to install (Piece of evidence 1.4) and the fact that smartphone are not trusted by their users (Piece of evidence 3.2, 3.3, 3.4). However, all the other observations and interpretations are new, and they represent a more deep and detailed understanding of what are the possible considerations that users formulate when installing new apps. This allow us to sketch some possible implication for the stakeholders of mobile phones.

B. Implications

The analysis of the observations and their interpretations highlights possible implications for the main actors of the app ecosystem. Among them, we considered users, developers, vendors and researchers.

Concerning app users, some discrepancies emerged between how users perceive privacy and security risks and their actual behaviour. For example, participant P4 claims not to care about exposing her data on social networks, but indeed she avoids using or installing apps that use or expose sensitive data. On the other side, participant P1 claims to be very scared about malicious apps that could misuse her data, even if she actually adopts some behaviours that can expose her to possible threats, such as using an unofficial app-store. Users should be informed more clearly and educated to understand security/privacy risks connected to their actions, possibility with the adoption a revised permission mechanism.

An interesting aspect, for requirements engineers, developers and vendors, concerns the opportunity to use flexible permission-based security models that take into account different user profiles and the changing context of use. This should lower the barrier towards the installation and use of apps, since the trade-off between perceived needs and privacy/security risks appears to be context dependent.

Finally, the adopted observational procedure seems effective in identifying peculiar characteristics of groups of users showing different behaviours with respect to privacy and security risks. This can be helpful for recognising different types of *persona* in user-centered design approaches [8]. This is in the line with the work of Aoyama [1], which exploited *persona* to identify key players in mobile services in a complex social environment. In addition to what described in [1], we observed that for mobile services and the more recent app ecosystem the aspect of *context* variability needs to be taken into account in the description of the characteristics of a *persona*. This is important to account for different behaviours of the same individual (and corresponding group) in different situations of mobile applications and service usage (see Observations 2.3 and 2.4).

C. Threats to Validity

In the following we discuss the threat to the validity of our results [17], with respect to the conclusion, internal, construct and external validity.

Conclusion validity threats concern issues that effect the ability to draw the correct conclusion on the observed phenomenon. For an observational study, they mainly deal with the risk of the researcher influencing the results (i.e., fishing) and the reliability of the measures. In fact, the experimenter could have assumptions and expectations, so she/he could record those observations that confirm such expectations and overlook those that are not expected. We try to limit this threat as much as possible by adopting best practices and guidelines [9], such as involving two independent observers on each session, considering two data sources (actions and interviews) and, eventually, asking some participants to confirm our findings.

Internal validity threats concern additional factors that may affect an observed variables. Participants could, in fact, react differently as time passes either positively (learning effect) or negatively (fatigue effects). We try to limit this threat by starting each session with the most critical part, i.e. the observation. The most boring part of the session, i.e. the profiling questionnaire, is placed at the end of the session. Moreover, to avoid participants guessing the experimental hypotheses, we never mention terms as *security*, *privacy*, *trade-off*, *value*, *decision-making* in the introduction and in the explanation of the study, and we ask all the security/privacy-related questions as last ones, to avoid any bias in the other answers.

Construct validity threats concern the relationship between theory and observation. In our case, they are mainly due to potential evaluation apprehension and participants bias. We limit these threats by explicitly stating that there are not correct/wrong answers and that participants will be not evaluated based on their results. Moreover, we assure participants that data will be kept confidential and will not be disclosed. Participants might be biased by their professional interests and their answers may come from their knowledge in the field of apps engineering, rather than from their actual experience as apps users. In order to mitigate this threat, we carefully designed the questionnaire by recalling concrete usage experiences, avoiding to use words like *security* and *privacy*. While starting a session, we generally state that we are interested in knowing how participants use their smartphones.

External validity threats concern the generalisation of the findings outside the experiment settings. Though only replications with other participants can confirm our findings, we try to limit this threat by involving people with different age, different hobbies and different degree of exposition to the social networks.

V. Related Works

In the domain of the smartphone apps, to the best of our knowledge the work by Chin et al. [7] is the most relevant, although with different objectives and investigation method. Chin et al. conducted a questionnaire-based survey on mobile phones usage, with the aim to investigate users perception of security and privacy risks, by comparing participants use of smartphone and of laptop computer. Some of their findings, namely the role of advice by friends in the selection on what app to install (Piece of evidence 1.4) and the fact that smartphone are not trusted by their users (Piece of evidence 3.2, 3.3, 3.4), were confirmed by our study. Differently from us, they directly asked questions to participants, so they pulled information from participants, rather than making information emerge from the observation, with the possibility that a question might have influenced the answers. Eventually, Chin et al. identified a limitation of the questionnaire as the possibility for the participants to elaborate on their answers to please the interviewer, and they claimed that an observational study would be appropriate to complement their survey. However, our study supports new observations and interpretations, as pointed out in Section IV, which contribute to a deeper and more detailed understanding of what are the possible considerations that users formulate when installing new apps. This allow us to argue about further implications for the stakeholders of mobile phones.

Worth mentioning is also earlier empirical studies guided by the Technology Acceptance Model (TAM) model that aim at finding empirical evidences on the relations betwen user perceived usefulness, ease of use, and system use with reference to cultural, social and economical aspects of a population of users. In a study conducted on a group of 579 Finnish smartphone users over a one-two months period in 2007, Verkasalo et al. [16] investigated the intention to use, or not to use, a smartphone-enabled service, from a predefined set. The participants were asked to install a logger in their smartphones. The logger was intended to run in the background to log the phone usage. Data are daily transmitted to a server for analysis purpose. Participants were then asked to fill in a web-questionnaire.

Similarly, Bouwman et al. [5] conducted a survey on a sample of 542 participants, representative of the Dutch consumers population in 2008, exploring the relationship between lifestyle traits, social influence, people's attitudes towards mobile innovations and the adoption of various types of mobile services. This type of survey, although powerful and statistically relevant, are mainly tailored to study the influence of cultural and social aspects on technology adoption, rather than to investigate deeper decision making mechanism that can help identifying how to better align user needs and technology.

VI. Conclusion

In this paper we presented the design and validation through a feasibility study of an observational study, aiming at exploring how smartphone users assess the trade-off between benefits and risks of the apps that they select and install. The results show that the experimental design is adequate to achieve the study objective and to support a set of preliminary observations. Replicating this study over large number of participants would allow us to consolidate the statistical relevance of the results', at the cost of a considerable investigation effort. With the current experiment approach we estimated about 5.5 person/hour overall for each interview, which includes 1 - 1.5 hour by the participant (interviewee), 1 hour by the observer and interviewer for the interview and 1 hour for the post-interview debriefing.

Concerning the results of the feasibility study presented in this paper, even if their statistical relevance is limited, the emerged observations are quite interesting and demonstrate that a structured observational study can be an effective approach to collect insights on this novel and largely unknown phenomenon. As future work we intend to build on these preliminary observations and identify what it makes sense to measure in a (more quantitative) experimental study. We also plan to experiment with a different investigation approach, i.e. asking participants to fill on-line questionnaires. A benefit will be the increase in the number of participants, limits will be represented by the less controllable experimental environment.

Finally, we aim at exploiting the proposed observational method in the context of software development processes, in particular, to assess the perception of technical and business risks in the activities of selection and adoption of open source software components.

Acknowledgement

This work is a result of the RISCOSS project, funded by the EC 7th Framework Programme FP7/2007-2013 under the agreement number 318249.

References

[1] M. Aoyama. Persona-and-scenario based requirements engineering for software embedded in digital consumer products. In *RE*, pages 85–94. IEEE Computer Society, 2005.

[2] E. R. Babbie. *The practice of social research.* Wadsworth Pub. Company, 2012.

[3] A. Begel, J. Bosch, and M.-A. D. Storey. Bridging software communities through social networking. *IEEE Software*, 30(1):26–28, 2013.

[4] A. Bhattacherjee. *Social Science Research: Principles, Methods, and Practices*. USF Tampa Bay Open Access Textbooks Collection, 2012.

[5] H. Bouwman, C. López-Nicolás, F. J. Molina-Castillo, and P. van Hattum. Consumer lifestyles: alternative adoption patterns for advanced mobile services. *IJMC*, 10(2):169–189, 2012.

[6] M. Ceccato, A. Marchetto, L. Mariani, C. D. Nguyen, and P. Tonella. An empirical study about the effectiveness of debugging when random test cases are used. In *34th International Conference on Software Engineering (ICSE), 2012*, pages 452–462. IEEE, 2012.

[7] E. Chin, A. P. Felt, V. Sekar, and D. Wagner. Measuring user confidence in smartphone security and privacy. In *Proceedings of the Eighth Symposium on Usable Privacy and Security*, page 1. ACM, 2012.

[8] A. Cooper. *The Inmates Are Running the Asylum.* SAMS, 1999.

[9] J. W. Creswell. *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage Publications, Incorporated, 2013.

[10] W. Enck. Defending users against smartphone apps: Techniques and future directions. In S. Jajodia and C. Mazumdar, editors, *ICISS*, volume 7093 of *Lecture Notes in Computer Science*, pages 49–70. Springer, 2011.

[11] X. Franch, A. Perini, H. H. Pibernat, and N. Seyff. Mobile technologies to enable users' informed decisions. In J. Cordeiro, D. A. Marca, and M. van Sinderen, editors, *ICSOFT*, pages 345–353. SciTePress, 2013.

[12] A. Gustafsson, A. Herrmann, and F. Huber. *Conjoint measurement [electronic resource]: methods and applications.* Springer, 2007.

[13] T. Roehm, R. Tiarks, R. Koschke, and W. Maalej. How do professional developers comprehend software? In *Proceedings of the 2012 International Conf. on Software Engineering*, ICSE 2012, pages 255–265, Piscataway, NJ, USA, 2012. IEEE Press.

[14] N. Seyff, G. Ollmann, and M. Bortenschlager. irequire: Gathering end-user requirements for new apps. In *RE*, pages 347–348. IEEE, 2011.

[15] A. Sutcliffe and P. Sawyer. Requirements Elicitation: Towards the Unknown Unknowns. In *RE 2013, 21st IEEE International Requirements Engineering Conference, Rio de Janeiro, Brasil, July 15?19, 2013*, pages 92–104. IEEE Computer Society, 2013.

[16] H. Verkasalo, C. López-Nicolás, F. J. Molina-Castillo, and H. Bouwman. Analysis of users and non-users of smartphone applications. *Telematics and Informatics*, 27(3):242–255, 2010.

[17] C. Wohlin, P. Runeson, M. Höst, M. Ohlsson, B. Regnell, and A. Wesslén. *Experimentation in Software Engineering - An Introduction*. Kluwer Academic Pub., 2000.