



<http://www.diva-portal.org>

Postprint

This is the accepted version of a paper presented at *The 11th Nordic ACM Conference on Human-Computer Interaction (NordiCHI '20)*.

Citation for the original published paper:

Bock, S., Momen, N. (2020)

Nudging the User with Privacy Indicator: A Study on the App Selection Behavior of the User

In: *Proceedings of the 11th Nordic ACM Conference on Human-Computer Interaction (NordiCHI '20)*, 60 (pp. 1-12). Tallinn, Estonia: ACM Digital Library

<https://doi.org/10.1145/3419249.3420111>

N.B. When citing this work, cite the original published paper.

© ACM 2020. This is the author's version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in {NordiCHI '20: Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences}, <https://doi.org/10.1145/3419249.3420111>.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:kau:diva-79307>

Nudging the User with Privacy Indicator: A Study on the App Selection Behavior of the User

SVEN BOCK*, Technische Universität Berlin, Germany

NURUL MOMEN*, Karlstad University, Sweden

This paper presents an empirical study on user behavior, decision making, and perception about privacy concern while selecting apps. An app store demo was presented to the user with a minor modification—a privacy indicator for each app. After carrying out several tasks using this modified mobile interface, participants were interviewed to document reasons behind their decisions, thought process, and perception regarding individual privacy. A total of 82 adults volunteered under the pretext of a usability study. A significant influence of the privacy indicator on their app selection behavior was observed, although this influence decreased in case of familiar apps. Furthermore, responses from questionnaires, data from eye-tracking device and documented interviews, with video confrontation showed coherence with respect to the corresponding app selection behavior.

CCS Concepts: • **Security and privacy** → **Privacy protections**; *Usability in security and privacy*; • **Human-centered computing** → Empirical studies in ubiquitous and mobile computing.

Additional Key Words and Phrases: Privacy indicator, Transparency, Decision making, User study.

ACM Reference Format:

Sven Bock and Nurul Momen. 2018. Nudging the User with Privacy Indicator: A Study on the App Selection Behavior of the User. *Proc. ACM Meas. Anal. Comput. Syst.* 37, 4, Article 111 (August 2018), 17 pages. <https://doi.org/10.1145/1122445.1122456>

1 INTRODUCTION

Today's app market offers popularity-based ranking which is entirely dependent on crowd-sourced user opinion. Convenience, ease of use, rich features and functionality are the established criteria for rating an app [21]. The excessive data harvesting nature of apps hardly features in the rating procedure [20]. As this negative attribute is rather kept hidden from the interface, apps' privacy invasive behavior has a very narrow scope to play an important part in a users' decision-making process while selecting an app. Users face hindrances to be aware of and to take preventive measures because of poor means to observe and to assess the consequences of data disclosure. Though users can solve numerous daily-life problems through finding a convenient app from online stores, their decisions to grant access to personal data could result to privacy implications [3, 18, 49]. Thus, informed decision-making about privacy by an ordinary user is hard to come by. This paper addresses the shortcoming and seeks for introducing a privacy indicator within app selection scenarios. We chose the Android app market for our study due to the open source nature of the platform and implementation feasibility.

*Both authors contributed equally to this research.

Authors' addresses: Sven Bock, sven.bock@mms.tu-berlin.de, Technische Universität Berlin, Germany; Nurul Momen, Karlstad University, Sweden, nurul.momen@kau.se.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2018 Association for Computing Machinery.

Manuscript submitted to ACM

Manuscript submitted to ACM

In order to address the aforementioned issue, we hypothesize that providing privacy-indicating cues in the app-store, could help the user in making informed decisions. The main research objective of this study is to analyze the impact of such *ex-ante*¹ cues [40] on the decision making process of users and to address the following questions: (a) Is the privacy indicator able to lead to a better judgment regarding the apps' trustworthiness and to provide the ease of selection? (b) Is there any significant difference in decision making behavior while selecting an app for a certain task? and (c) Is the indicator an adequate tool to illustrate the intrusiveness of apps in terms of privacy?

This paper presents our design, implementation and evaluation processes of an *indicator* as an *ex-ante privacy indicating cue* for the app-store. Our contributions are: (i) Through an online study and an empirical study that included 82 participants, we show that the impact of privacy indicators on app selection behavior has statistical significance. (ii) By varying the degrees of participatory background information and introduction, we show that the mere presence of the indicator caused participants to select more privacy-friendly apps for six out of eight app categories. (iii) For Messaging and Video-call categories, the indicator had limited or no impact. We suspect, this is because of background preferences for very familiar apps, suggesting that privacy concern decreases with respect to app familiarity. (iv) Eye-tracking data yielded gaze-behavior of the participants, which indicates that time requirement for decision-making remains insignificant. Hence, the indicator can be considered as easy to perceive and does not bring burdensome responsibility for the user.

The rest of this article is organized as following: Section 2 provides a brief overview of prior research works conducted within this area. Section 3 describes how we designed indicators, conducted the online study and selected an indicator that is easy to perceive. Section 4 documents how the empirical study was performed along with the results collected from it. Section 5 elaborates on interpretations, and evaluation of results. We discuss the limitations of the findings and our future research plans in Section 6. We draw a conclusion of this paper in Section 7.

2 BACKGROUND

There are numerous prior research works on the security and privacy aspects of Android apps. This section outlines the related work on apps' security and privacy issues, efforts to aid the user to protect privacy and users' decision dilemma. Studies were also conducted to report on users' awareness and concerns regarding privacy, identifying privacy leaks, the usability challenges of privacy controls. Hence, many solutions were introduced to aid users in managing privacy. We briefly discuss the relevant background to formulate the problem and hypothesis.

2.1 App behavior and Privacy Friendliness

The Android operating system relies on a permission-based access control model that in place to guard user data and sensors [13]. Depending on type of permissions, approval from the user is required during the first use of the app for granting access to resources [12]. As the platform offers binary choices (accept / decline), it is difficult to perceive consequence for granting access and assess the risk, if not impossible. Moreover, information is hardly available about the usage of user data once access has been granted. So, a white-card privilege to the available system resource is given to the app that leaves access decisions about sensitive personal data to arbitrary programs and services. This is a problem for the data subject who suffers the lack of appropriate information to make decisions regarding privacy preferences. Several prior works pointed out this problem that mostly put emphasis on apps' data access potential, frequency,

¹Before the event (data disclosure) takes place.

consequences and privacy implications which can be identified as parameters to define *apps' privacy-friendliness* [16, 17, 20, 22, 35, 37–39, 50].

We intend to consider the outcome of app behavior analysis as the input to our indicator ranking mechanism. For instance, Hatamian et al. [20] measured app behavior based on four parameters: data access potential (permission requirement), frequency of resource usage (runtime access), regulatory (GDPR—General Data Protection Regulation [9]) compliance according to corresponding privacy policy and user review analysis. Before including a comprehensive privacy score generating mechanism into our research, the prime goal is to find out whether a privacy cue can make a difference in user decisions within this evolved context or not. However, indicator ranking mechanism remains out of scope for this article.

2.2 Related Work

Though a decent amount of research effort has been invested to study app behavior, the app market is mostly inclined with ratings that are harvested from crowd-sourced user feedback. So, apps' data access potential is not easy to unearth and the user usually struggles to deal with individual privacy while deciding upon installing mobile apps [1, 5, 27]. There exists an array of prior works that shed light on users' remorse and struggle with understanding privacy risks associated with apps [43, 45, 47, 50] and offer solutions to aid the user [2, 31, 32, 36, 46].

Rajivan and Camp used visual cues to support the user in decision making before giving consent to requested permissions [42]. Kraus et al. used a permission scale to communicate potential risks to the user [28]. Gu et al. concluded that a high level of malaise is perceived by users when an app requires accesses to rights (perceived permission sensitivity) [19]. This raises users' privacy concerns while downloading apps. In contrast, the apps' familiarity reduces these concerns and the effect of justifying access rights on privacy concerns is heavily dependent on previous bad experiences that the user has faced regarding privacy with mobile devices.

Kelley et al. found that certain contextual notices, such as privacy symbols, can be effectively used to influence the subsequent use of privacy [26]. In [27], Kelley et al. performed another study which contained 20-participant lab study and a 366-participant online experiment. In this study, participants were presented with a short display about privacy facts that included a brief description about potential consequences. However, all these privacy cues from [25–28, 42] are addressing secondary user interface—app details to communicate privacy risks. It is possible to see the privacy cue after selecting an app which implies to decisions being made without cues' influence. Once the user has selected an app and become aware of the warning from privacy cue, it is cumbersome to go back to the app selection; although the user is aware of the aggressiveness of the app. Therefore, these studies are not supporting decision-making comparison in primary interface—app store, and they require more cognitive effort besides technical knowledge, e.g. permission.

Moreover, these studies were conducted in a different context compared to ours because Androids' access control model had gone through several changes and before evolving into the current one; most notable changes were the run-time permissions in 2016 [14] and restrictions on app-behavior in 2018 [11]. So, the user does not see permission requirements prior to installation because their consent is required during run-time. Thus, it can be deduced that privacy has a very narrow scope to play any part in decision-making process of the user while choosing an app for installation. These significant context changes require pursuing exploration with renewed challenges.

Understandably, users hardly pay attention to data disclosure and/or to permission requirements prior to app installation [4, 44]. However, users are more concerned about their privacy when they realize that their decisions have put them at risk of data leakage through third-party apps [15, 24, 48]. Also, revealing apps' data access and sharing

practice can upset the user [47] and thus, they seek for remedy [50]. So, we can infer that privacy-facts (that inherit complexity) under-perform against app-ratings (that come with simplicity) while users decide upon app installation.

Two significant research efforts can be noticed from prior works: (a) to develop a comprehensive mechanism to determine apps' privacy-friendliness and (b) to communicate privacy risks to the user through an easily perceivable illustration. We see the potential to fill the gap between them through this work. To address the challenge against overcoming complexity issues and support informed decision making in ex-ante scenarios, we put priority on designing easy-to-perceive indicators. To keep the effort of the user as low as possible when selecting an app, the indicator should bring a high degree of intuitiveness. Hence, we first elaborate on the design process of privacy indicators and then, evaluate them based on their cognitive appeal.

3 PRE-STUDY: INDICATOR SELECTION

In this section, we elaborate on the design process of five different variants of the privacy indicator. Then we describe how an online survey was run to choose an indicator that is easy to perceive, to understand and has the potential to cause awareness. The outperforming indicator was then implemented in a simulated app market for conducting a user study that is going to be discussed later.

3.1 Finding an Appealing Indicator

In the design phase, we addressed one question—*How can we construct an easily perceivable indicator with potential to yield spontaneous awareness?* For a comparative analysis, five indicator instances were chosen to demonstrate different levels of privacy risks associated with apps which are shown in Table 1. Our indicator designing process is elaborated in this section.

3.1.1 Designing Indicators. In [25], Kelley et al. presented privacy labels with detailed privacy facts, similar to nutrition facts on labels of food produce. We argue that complex and multi-dimensional information could compel the user to ignore. So, a goal was set to design an unidirectional information based and cognitively easy to perceive indicator. In this design process, color is considered as a critical attribute for accurate identification, while other attributes such as size, brightness, and shape can vary without affecting identification [7]. In [8], Cimbalo et al. conducted a study of children and students in terms of emotionally-colored images and color choices. Green is always attributed to happy and therefore positive emotions, while red is more likely associated sad and negative emotions. In [6], Benbasat et al. conducted a study that classified information as more understandable and better in terms of decision-making when it was presented in multiple colors compared to monochrome. Thus, for our design, we adopted subdivision of the indicator into five different color sections from dark-green to light-green, yellow, and orange to red. The indicator should require little cognitive effort to understand it, so that the information provided can be better taken into account in decision-making. The visual cues, in this case, the color-coded symbols of the privacy indicator, contribute to informed decisions as previously demonstrated by Hibbard and Peters in [23].

Having decided on a color-coded indicator, the next task was to find the right shape for it. An extensive literature research unfortunately did not yield any further information about an optimized indicator shape. Due to this reason, a focus group discussion (participated by several researchers from two research groups) yielded indicators that can also be found in everyday life and that are already familiar to the users: (a) Arrow-scale bar from the thermometer, (b) Label bar from the product test, (c) Arrow-scale meter from the speedometer, (d) Smiles from the evaluation of customer satisfaction in retail, and (e) Bubbles from a traffic light. Both a and c show range of the scale and point towards the

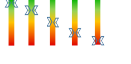




Indicator	Appearance	N	Min	Max	Average	SD
Arrow-scale bar		168	1	10	6.64	2.66
Label bar		168	1	10	8.25	2.28
Arrow-scale meter		168	1	10	7.80	2.45
Smiles		168	1	10	7.64	2.50
Bubbles		168	1	10	7.02	2.68

Table 1. Statistical description of survey on different privacy indicators. Respondents (N=168) could rate the indicators between 1–10. *Label bar* stood out among the rest, based on calculated average and standard deviation (SD).

indicating instance. On the other hand, *d* and *e* visualizes the indicating instance. Only *b* shows the instances with written descriptions.

Furthermore, we decided on a five-point scale, because of the cognitive limitation of humans to process information. In [34], Miller states that a human being can keep 7 ± 2 units of information (chunks) in the short-term memory. This ability is genetic with individual exceptions and cannot be increased by training. We took the lower limit of chunks that can be stored in short-term memory, so that the majority of end-users can use the indicator with low cognitive effort. In addition, we opted for an odd-numbered scale so that apps can be labeled as *moderate* besides *secure* or *insecure*.

3.1.2 Online Survey. To verify which design is the most appropriate one, an online survey was conducted. The participants were presented with app store demo as shown in Fig. 1, but with five variants of indicator as shown in Fig. 1. They were asked to rate each scale from 1 to 10 according to following attributes: (1) *Discriminability*—Are the colors discriminable? (2) *Ambiguity*—Is the indicator/ graphic unambiguously/ clearly understandable? (3) *Readability*—How well readable is the information that is presented by the indicator/ graph? (4) *Comprehension*—Do you understand the information which is illustrated by the indicator/ graphic? (5) *Color (Appeal)*—Is the intensity of the color of the indicator/graphic appropriate? (6) *Size (Appeal)*—Is the size of the indicator/graphic appropriate?

Their rating scores were recorded to determine visual and cognitive appeal of the indicators. The survey was active from the end of August 2018 until the end of December 2018. During this time, a total of 55 individuals participated in the survey. Unfortunately, only 28 individuals completed the questionnaire with all five indicators. So, only the completed results are taken into account. The survey consisted of 28 people, who gave ratings in the six categories, and in the end, it resulted in 168 ratings per indicator.

3.2 Pre-study Results

The descriptive evaluation in Table 1 shows the following mean values, Standard Deviation (SD), in descending order for a span of one to ten: *Label bar* $M = 8.25$ (2.28), *Arrow scale bar* $M = 7.80$ (2.50), *Smiles* $M = 7.64$ (2.51), *Bubbles* $M = 7.02$ (2.68) and *Arrow-scale Bar* $M = 6.64$ (2.66). The data sets of the responses of each indicator were then tested for normal distribution using the Kolmogorov-Smirnov test [30]. The outcome was significant for all indicators ($p < .000$), so that a normal distribution can not be assumed. Then a non-parametric Friedman test [10] was used and it resulted in a significant difference ($\chi^2 = 54.35$; $p < .000$). In addition, we compared the individual groups of pairs with the help of the Wilcoxon test and found significant differences in all pairs with $p = .000$ apart from the pair Arrow Scale Bar and Bubbles with $p = .188$.

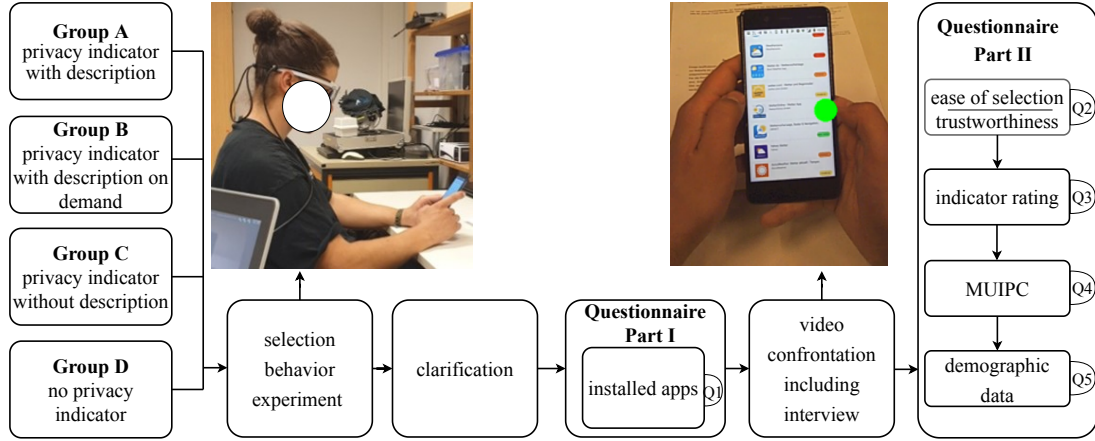


Fig. 1. An overview of the lab study procedure.

Based on the analysis of ratings from respondents, a favoritism for the *Label Bar* could be recognized. The Friedman test showed a significant difference between the scores of each scale, so a decision regarding implementation was made in favor of the *Label Bar*. Justifications for the assessments were not included in the study. However, there is reason to believe that the *Label Bar* can be interpreted with less cognitive effort than the other indicators and therefore, has received better rating. It can also be considered as achromatopsia-friendly.

4 LAB-STUDY: USER PREFERENCE

This section elaborates the lab-study to illustrate user preference towards privacy indicator on app store. First, we discuss the study's work-flow and the corresponding parameters that we measured. Second, we describe the results, highlighting on the app selection behavior of participants. We also reflect on the statistical significance of our findings and feedback from participants regarding the privacy-indicator-embedded app store.

4.1 Methodology: Empirical Study Procedure

To evaluate the outcome of the pre-study, we implemented the *Label bar* in an app-market prototype. As shown in Fig. 1, the *Label bar* was placed right next to the app logo, so that the user could view and compare the list of apps along with their data access potential and make a decision on the primary interface without needing to explore details of every app. Our goal was to determine users' app selection behavior during given scenarios. The testees were asked to fulfill different tasks using several provided apps and then interviewed to explain their actions. The study design was approved by the Ethics Committee with appropriate authority.

4.1.1 Recruitment of participants and Demography. Advertisement for our study was circulated through online platforms which included blogs, social media, etc. as well as offline means that included posters hanging on notice boards of several institutions located within reachable proximity. Due to the fact that participants need to be physically present in the lab, most of them are recruited from our geographically convenient location. Nonetheless, the majority of the participants have a background that is diverse due to the cosmopolitan nature of a big city. A total of 82 German-speaking persons took part in the lab study. The demography of the population is illustrated in Fig. 2. Each subject was compensated

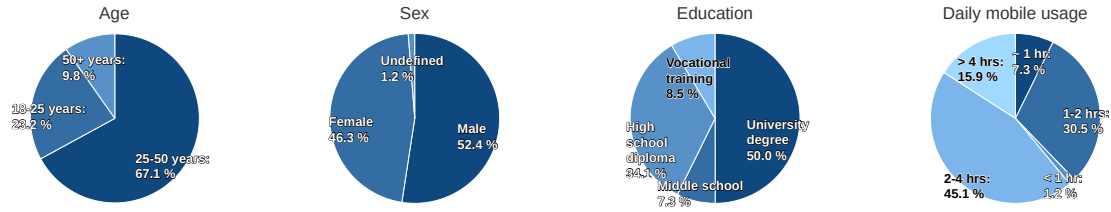


Fig. 2. Demography of participants in lab study.

with €10 for participating in the study. If the study lasted longer than one hour, the subject was compensated with an additional €2.50 per 15 minutes elapsed.

4.1.2 Distribution of participants. This empirical study was constructed with a 4 x 1 between-subjects design. The subjects were randomly assigned, considering a mix of age, sex, occupation, academic and geographical backgrounds, to one of the following four groups with different experimental conditions, based on the different availability of information in the app market. Thus, the subjects from one group only went through one of the four experimental conditions: For Group A, an app-store interface is presented with privacy indicator, along with a detailed description as an introduction. For Group B, an app-store interface is presented with privacy indicator, along with a detailed description on demand, by clicking on the privacy indicator. For Group C, an app-store interface is presented with privacy indicator, but without any detailed description. For Group D, an app-store interface is presented with no privacy indicator. This group represents the control group of this experiment. Our aim was to find out if the participants were paying attention to the privacy indicators, to analyze the decisions they make and to combine them with a questionnaire.

4.1.3 Scenarios and App-categories. The app-store demo consisted of eight categories, and they were associated with mock up scenarios. Each category had ten apps and their ratings were removed due to possibility of interference in decision-making. All the participants had to go through each of the scenarios. The scenarios and app-categories are described below:

- (1) Messenger: Imagine you want to schedule a dinner with a friend. Write her a short message.
- (2) Video calling: Imagine you are on a trip and want to talk to a friend. Start a video call with her.
- (3) Weather: Imagine you are packing your suitcase for a trip and want to look at the weather forecast to pack the right things.
- (4) E-mail: Imagine you want to invite a friend to a concert. Write her an e-mail.
- (5) Music: Imagine you want to hear a song from the current charts on your smartphone.
- (6) Fitness: Imagine that you want to track your physical activities. Set up a fitness app for it.
- (7) Games: Imagine you are about to board on a long flight. For entertainment, you want to download the game Sudoku, which you can play on your smartphone.
- (8) News: Imagine you want to know the latest news. Open a news app.

4.1.4 Workflow of an Experimental-session in Lab. Participants are left unaware of the purpose of the study. They decided to join the study under the pretext to rate the usability of an app-store. Upon appointment confirmation, they are randomly assigned to group A, B, C, and D. Figure 1 shows the lab environment. To accelerate the study for each and every app, an account was already created, if needed. Upon completion of all tasks, they were interviewed and

had to complete a questionnaire to examine acceptance, trustworthiness, and usability. To identify the interference in decision-making, they were asked if the apps used in the experiment were already installed in their private devices. The steps of the lab-session are illustrated by the flow chart in Fig. 1.

4.2 Empirical Results: Users' App Selection behavior

We assumed that implementing a privacy indicator in an app-store would result in a more appropriate selection of apps to protect privacy. To compare the selection behavior with respect to the indicator illustrated beside the app, we assigned each app selection a rating (ranging from 1 to 5; considering 1 as very critical and 5 as very safe). The mean values which could be seen as a privacy score in Table 2, show a tendency of a higher value for the groups (A, B, and C) with indicator compared to the group (D) without indicator. This primarily confirms the assumption that the subjects who had illustrated an indicator selected an app that was classified as less critical.

In order to apply the appropriate statistical tests for each of the data sets, the sets were tested for normal distribution using the Kolmogorov-Smirnoff [30] test and either parametric or non-parametric methods were applied.

4.2.1 Privacy Friendliness of the Selection behavior. An overview of the users' app-selection behavior from statistical analysis is presented in Table 2. The test showed that the data sets of the privacy score deviated significantly from a normal distribution, so that the non-parametric Kruskal Wallis test [29] for more than two independent samples was performed. A significant difference regarding the selection behavior among all four groups resulted in the following app categories: (1) Weather ($\chi^2 = 13.952$; $p < .003$), (2) E-Mail ($\chi^2 = 11.87$; $p < .008$), (3) Music ($\chi^2 = 18.862$; $p < .000$), (4) Fitness ($\chi^2 = 18.649$; $p < .000$), (5) Games ($\chi^2 = 20.9$; $p < .000$) and (6) News ($\chi^2 = 20.926$; $p < .000$). Additionally, Bonferroni-correction was carried out to counteract the problem of multiple comparisons. The corrected p-Value now equals .0062. After the correction, only the category (2) E-Mail must be excluded from the significant results. Further, no significant difference could be determined for the outcomes of the Messenger and Weather app categories. Subsequently, the first three groups were combined into one group to perform a Mann-Whitney U test [33], which examines two distributions for significant differences. The test showed significant differences in the same categories that were considered significant in the Kruskal Wallis test before the Bonferroni-correction [29]. In addition, the individual groups were examined for significant differences with each other using the Mann-Whitney-U test [33]. While comparing the following pairs of groups, significant difference ($p < .05$) was found in following cases:

Group A and B: (i) Video-call ($U = 139$; $z = -2.092$), (ii) E-Mail ($U = 132$; $z = -1.92$) and (iii) Music ($U = 93.5$; $z = -3.616$).

Group A and C: (i) News ($U = 169$; $z = -2.421$).

Group A and D: (i) Weather ($U = 69$; $z = -3.637$), (ii) E-Mail ($U = 76$; $z = -3.705$), (iii) Fitness ($U = 93.5$; $z = -3.251$), (iv) Games ($U = 75.5$; $z = -3.6$), and (v) News ($U = 73$; $z = -4.055$).

Group B and D: (i) Weather ($U = 94.5$; $z = -2.602$), (ii) Music ($U = 81.5$; $z = -3.494$), (iii) Fitness ($U = 73$; $z = -3.504$), (iv) Games ($U = 63.5$; $z = -3.683$) and (v) News ($U = 81$; $z = -3.259$).

Group B and C: only Music ($U = 131.5$; $z = -2.524$).

Group C and D: (i) Fitness ($U = 132.5$; $z = -2.389$) and (ii) Games ($U = 90.5$; $z = -3.527$).

Notably, in case of comparisons concerning (2) and (3), they are the same app categories found significant in the Kruskal Wallis test[29], which compares every group with each other.

4.2.2 Ease of App Selection and Apps' Trustworthiness. In Q2, a subset of Questionnaire part II, shown in Fig. 1, participants provided feedback regarding the ease of *app selection* and *trustworthiness* of the app (on a scale of 1 =

Groups		Messenger	Video-call	Weather	E-mail	Music	Fitness	Games	News
A: indicator with detailed description & explanation	Mean	3.10	4.24	4.00	4.10	4.05	4.71	4.38	4.86
	N	21	21	21	21	21	21	21	21
	Std. Dev.	0.94	0.54	1.00	0.94	0.22	0.64	1.07	0.65
B: indicator with description & explanation on-demand	Mean	3.00	3.79	3.58	3.32	4.58	4.83	4.47	4.58
	N	19	19	19	19	19	19	19	19
	Std. Dev.	1.05	0.79	1.02	1.34	0.51	0.71	1.07	1.07
C: indicator without description or explanation	Mean	2.61	4.04	3.39	3.43	4.00	4.30	4.35	3.96
	N	23	23	23	23	23	23	23	23
	Std. Dev.	1.08	0.93	1.27	1.31	0.90	1.18	1.11	1.52
D: app store without any indicator	Mean	2.95	3.95	2.63	3.00	3.89	3.21	3.05	3.00
	N	19	19	19	19	19	19	19	19
	Std. Dev.	0.71	0.62	0.96	0.67	0.74	1.58	1.03	1.49

Table 2. Statistical significance of participants' app selection behavior: Highest and lowest mean values amongst the four groups per category are colored in green and red, respectively.

strongly agree, to 7 = strongly disagree). The collected data indicates that the highest averages, with the exception of ease of selection in Messenger category, are all to be found in Group A, B, and C. Thus, it could be assumed that with the exception of the Messenger category, the indicator has led the participants to find it easier to choose an appropriate app to accomplish the task and rated the app as more trustworthy than the participants without indicator. To test this assumption, a Kruskal Wallis test [29] was performed due to the absence of a normal distribution. This resulted into findings with significant differences in terms of trustworthiness of the app in the categories (1) E-Mail ($\chi^2 = 8.488$; $p < .037$) and (2) Fitness ($\chi^2 = 12.563$; $p < .006$).

For the ease of app selection, significant difference in the averages could be found for Fitness ($\chi^2 = 12.133$; $p < .007$) category. But after carrying out a Bonferroni-correction no significance $p < .003$ was found. In addition, we compared the individual groups of pairs with the help of the Mann-Whitney-U Test and were not able to find a significant difference comparing all pairs in Group A and B. While comparing the following pairs of groups, a significant difference ($p < .005$) was found in following cases:

- Group A and C: (i) Weather ($U = 154$; $z = -2.133$), (ii) E-Mail ($U = 129.5$; $z = -2.252$), (iii) Music ($U = 152$; $z = -2.289$), (iv) Fitness ($U = 94.5$; $z = -3.007$) and (v) Games ($U = 127$; $z = -2.582$) for ease of selection as well as (i) E-Mail ($U = 127.5$; $z = -2.386$) and (ii) Games ($U = 127.5$; $z = -2.574$) for trustworthiness.
- Group A and D: (i) Fitness for ease of selection ($U = 58$; $z = -3.034$) and trustworthiness ($U = 65.5$; $z = -2.574$) as well as (ii) Video-call ($U = 122$; $z = -2.145$) and (iii) E-Mail ($U = 93.5$; $z = -2.59$) for trustworthiness.
- Group B and C: (i) Weather ($U = 125$; $z = -2.446$) for ease of selection and trustworthiness ($U = 135.5$; $z = -2.179$) as well as (ii) Fitness ($U = 132$; $z = -2.013$) for trustworthiness and (iii) News ($U = 140.5$; $z = -2.068$) for ease of selection and for trustworthiness ($U = 128$; $z = -2.392$).
- Group B and D: (i) Messenger ($U = 106.5$; $z = -2.193$), (ii) Video-call ($U = 112$; $z = -2.059$) and (iii) Fitness ($U = 57.5$; $z = -3.029$) for trustworthiness.

No significant difference can be found while comparing Group C with Group D.

4.2.3 Indicator Evaluation by Participants. In Q3, a subset of Questionnaire part II, shown in Fig. 1, participants provided evaluation ratings, based on *indicators' trustworthiness, usability, understanding, intuitiveness, validity, reliability, impact,*

feeling of protection, demand, ambiguity, lack of information, complexity, willingness to use, apps' trustworthiness. It is noticeable that on a scale of 1 = strongly agree, to 7 = strongly disagree, overall a positive rating was given to positive statements. Overall, the negative statements (ambiguity, lack of information, complexity) were disagreed by the majority.

In order to archive the subjective opinion of the testee, we did not define in detail how to interpret the given terms. Comparing the groups with each other, a tendency for better rating in group A can be found. Group C gave the worst scores, while group B's evaluations are mostly found between group A and C. A normal distribution was precluded for all value pairs, so the non-parametric Kruskal Wallis test was applied [29], which, however, could only show a significant difference for the lack of information ($\chi^2 = 8.583$; $p < .014$). A pairwise analyses between the groups A, B and C using the Mann-Whitney-U test confirmed a significant difference between group A and C regarding the impact on the selection behavior and lack of information. The latter is also significant between B and C.

4.2.4 MUIPC—Mobile Users' Information Privacy Concerns. In Q4, a subset of Questionnaire part II, shown in Fig. 1, the standardized MUIPC [51] questionnaire was used to capture the current concerns regarding data protection in the area of mobile devices and thus clarify the necessity of the indicator. By analyzing responses of participants, we can see that a surprisingly high percentage of the participants agreed with the given statements. In total 84.2% believed that the location of their mobile device is at least temporarily monitored. 73.3% are worried that apps are collecting too much information about them. 68.3% did say that they are worried that apps monitor their activity on their mobile device. 68.3% do feel like others know more about them through their use of apps than they like. 73.2% believe that because of their use of apps, information about them that they consider private is now better available to others than they would like. 73.2% feel that because of their use of apps, information about them is available that, when used, invades their privacy. 69.5% are concerned that apps may use their personal information for other purposes without notifying them or soliciting their permission. 76.8% are concerned, that if they provide personal information about the use of apps, that the apps will use their information for other purposes. 47.6% are worried that apps can share their personal information with other instances without soliciting their permission. Lastly, 74.3% will probably disclose their personal information to use apps in the next 12 months.

4.2.5 Relationship between app selection and privacy concerns (MUIPC). In order to determine the correlation between the privacy scores of the selected apps and the result of the questionnaire to measure the concerns of mobile phone users regarding data protection, correlations between the scores and the 13 items of the MUIPC were checked. For the data sets, which showed a normal distribution in the privacy score as well as in the MUIPC and were correlated with each other, the Pearson correlation was used. For all other data sets the Spearman-Rho correlation [41] was used. Correlations with a significant effect are described below.

E-Mail: 'I have the feeling that others know more about me through my use of apps than I am comfortable with' ($r(80) = 0.245$; $p < .03$), 'I feel that due to my use of apps, information about me is available that, when used, invades my privacy' ($r(79) = 0.26$; $p < .02$), 'I am concerned that apps may use my personal information for other purposes without notifying me or obtaining my authorization' ($r(79) = 0.285$; $p < .01$), 'When I hand out personal information to use apps, I am concerned that the apps will use my information for other purposes' ($r(80) = 0.287$; $p < .01$) and 'I am concerned that apps may share my personal data with other instances without my authorization' ($r(79) = 0.285$; $p < .01$). All other correlation pairs resulted to be not significant.

Music: 'I believe that the location of my mobile device is monitored at least temporarily' ($r(80) = -0.313$; $p < .00$) and 'How many times have you personally experienced incidents within the last 2 years where your personal information was used by companies or e-commerce websites without your permission?' ($r(80) = -0.22$; $p < .05$). In both cases, a

negative correlation could be shown. It should be noted that the first test of site monitoring shows a medium effect level and the second test of the frequency of data use without consent shows a small effect level. Video-call: 'I have the feeling that others know more about me through my use of apps than I am comfortable with' ($r(80) = 0.225$; $p < .04$) correlated with a small effect. No significant correlation was found between the MUIPC results and the Weather, Games, News, and Messenger scenarios.

4.2.6 Comparison of Fixation Time from Gaze-behavior Data. Unfortunately, it was possible to analyze only 72 eye-tracking videos out of the 82 participants. In eight cases, participants' spectacles caused distortion to the eye-tracker pointer. In two cases the video files were damaged and could not be opened.

On an average, the participants looked at the indicator for 12.44 seconds ($SD = 10.29$), at the text for 44.16 seconds ($SD = 26.71$) and at the symbol of the app for 18.22 seconds ($SD = 19.29$). With regard to the indicator, when comparing the groups A ($M = 12.46$ seconds; $SD = 12.06$), B ($M = 12.41$; $SD = 10.28$) and C ($M = 12.44$; $SD = 8.49$), no significant difference is apparent at the first glance. The result from the Kruskal-Wallis test, as well as from the Mann-Whitney test did not show significant difference between the groups. The mean values of the fixation times on the symbol of the app in the app market show that group A has the longest fixation time of 22.1 ($SD = 16.55$) seconds, while group B has 16.25 seconds ($SD = 12.04$) and group C has 14.89 seconds ($SD = 20.32$). The participants in group D invested an average of 19.78 seconds ($SD = 25.6$) to look at the symbol and are therefore between the average values of group A and B. If we look at the fixation times of the text (app name), an inverse behavior can be observed. Group A took the shortest fixation time with 35.54 seconds ($SD = 25.27$), while group B with 44.47 seconds ($SD = 26.97$) and group C with 50.57 ($SD = 23.87$) seconds. However, after testing for significance using the Mann-Whitney U-test, only a significant difference with a medium effect was identified when comparing fixation times between group A and C with respect to the text ($U = 96.5$; $z = -2.441$; $p < .015$; $r = 0.4$).

4.2.7 Familiarity of the selected app. From the answers to Q1, Questionnaire Part-I in Fig. 1, it is noticeable that the highest values for 'installed and already in use' were found in case of Messenger (78.6%), Music (48.8%), and Video-call (39.3%); while the highest values for 'not yet installed and not in use' were found in case of Games (92.9%), Fitness (70.2%), News (56.0%), E-mail (44.0%) and Weather (29.0%). A closer look at the results sorted by groups regarding the familiarity of the selected app revealed a similar distribution, although of course there are some categories in the scenarios that differ between the groups, e.g. the category 'not installed and not used' in the Messenger scenario (Group A = 14.29%, Group B = 31.60%, Group C = 13.00%, Group D = 5.30%). These results also explain the partially different distributions of the selected apps within the scenarios. Some scenarios, e.g. the weather (Wetter-Online (26.8%), Wetter.com (25.6%), Wetter.de (22%) and Morecast (19.5%)) and the fitness ((39%) Pedometer- Free Pedometer & Calorie Meter and (30.5%) FitBit), have a more homogeneous distribution with regard to the choice of app, while other scenarios, e.g. that of video-calling (Skype (72%), Viber (17.1%)), music (Spotify (74.4%), Soundcloud (22%)) and messenger (Whatsapp (63.4%), Facebook Messenger (11%), Telegram (8.5%)) show one-sided distribution due to apps' familiarity.

5 DISCUSSION

In this section, we discuss the interpretation of results collected and analyzed. From Table 2, a comparative user awareness can be observed in the app selection behavior for all the given scenarios. From the analysis of collected data, a general observation can be made: *privacy concern decreases with app-familiarity*. Participants deliberately ignored the indicators due to apps' familiarity attribute. In Table 2, this observation is prominent in Messenger and Video-call category due to frequent selection of Whatsapp and Skype. However, skeptic and cautious app selection behavior, by

choosing safer options often if more privacy details provided, can be observed in case of Weather, Fitness and News apps. Our theory behind causing this phenomenon is that the large amount of press and media releases highlighting negative aspects of apps' personal data collection have raised awareness. Weather apps' location data gathering, Fitness apps' data collection resulting in revelation of army bases' map, and 'fake-news' becoming buzzword about misinformation, may have led to more privacy-aware app selection.

5.1 Assisting the User in App Selection

At first glance, we observed that, with the exception of the Messenger category, the mean values for ease of app selection were highest in the groups (A, B, and C) with indicator. When evaluating the trustworthiness of the app, it was possible to find the highest mean values in these groups. The results also indicate that (for the categories Video-call, Weather, E-mail, Fitness, and News) users having indicator assisted interface were significantly more adequate at evaluating the trustworthiness of the app than users who did not have the indicator available. Here, the results could again be explained by the familiarity of the app. Using the example of Skype, a familiar app in combination with a secure indicator leads to a higher rating in the trustworthiness, as an app without indicator. A familiar app with an insecure indicator, within a category like the messenger app WhatsApp, leads to a lower trustworthiness rating because the app is chosen for its familiarity despite the unsafe classification by the indicator, but the subject is aware of the insecurity of the app and thus indicates a lower trustworthiness. If the app is not known and the indicator categorizes it as safe, then the trustworthiness of the app is considered high. An unfamiliar app with an insecure classification by the indicator is usually not selected and thus has no influence on the rating. A familiar app without an indicator is heavily dependent on its reputation, which is evident from the collected data. A similar pattern can be seen in the case of ease of selection. There, only Weather and Fitness categories had a significantly higher mean. There was no favored app in these two categories, which led to a significantly higher ease of selection for the participants with an indicator.

5.2 Influencing App Choice through Nudging

Comparing the privacy-preserving app selection scores in Table 2, the highest means of the scores belong to Group A (Messenger, Video-call, Weather, and E-Mail) and to Group B (Music, Fitness, Games, and News). It should be noted that the lowest scores are always to be found in group D (without indicator), with an exception in the Messenger and Video-call categories. The Kruskal Wallis Test revealed significant differences in all app categories except Messenger and Video-call. Thus, the indicator had a significant impact on the selection behavior, leading to a selection of less invasive apps in Weather, Email, Music, Fitness, Games, and News categories. The lack of significance in comparison of Messenger and Video-call can be explained by the nullifying parameter—familiarity of the app. The interviews point that most participants chose an app they already knew. Also, the first part of the questionnaire denotes that most of the selected apps for the task were already known to the users, or are currently used by them. Furthermore, the results show that 63% of the participants chose WhatsApp and the remaining 37% were split between the other apps. Similar results can be found in the Video-call category: Skype was chosen by 74% and Viber by 17% of the participants. The remaining 9% were divided between the other apps. The reason for the lack of significant differences could be seen if we compare the values with those of the Weather category, which are clearly distributed more homogeneously (Wetter-Online = 24%, Wetter.com = 26%, Wetter.de = 22%, Morecast = 20%, Weatherzone = 2%, Bayer Agrar Wetter = 2%, Yr = 1%). Thus, it could be concluded that with increasing familiarity of the app, the influence of the indicator decreases.

Furthermore, the importance of description can be seen by comparing the group D with each of the individual groups: A, B and C. While in group A significant differences can be found in 6 categories and in group B in 5 categories, between

C and D only 2 app categories (Fitness and Games) can be found, whose mean values differ significantly. However, it should be noted that all averages of group A to C, except for the categories Messenger and Video-call, are higher than the averages of group D.

5.3 Comparison of Gaze-behavior

When comparing the glance time on the privacy indicator, based on the similar mean values between the Groups, it was already possible to assume that no significant difference to be found. This assumption was confirmed after the significance tests had been performed. Therefore, it can be deduced that the available information about the indicator has no impact on the observation time of the participant. However, it is noticeable that the observation time on the app-logo appears to be positively related to the available information about the indicator (i.e. the glance time increases with provided explanation). An inverse correlation can be observed among the data from different Groups (A, B, and C), regarding the glance time on the app-name (i.e. the glance time decreases with the provided explanation about the indicator). Here, it can be observed that the explanation of the privacy indicator led to higher attention at the app-logo and paying less attention to the app-name. A probable explanation for this behavior could be that the privacy indicator with explanation has a higher importance for the participants because they already know the function of the indicator and therefore, include it in their decision-making process.

Similarly, less attention is paid to the app-name in Group A compared to the other Groups, because the participants in this Group had another tool at their disposal to assess the apps. This may have led the participant to spend less time reading the app-name and more time looking at the app-logo. However, a significant difference was found only between Group A and C regarding the glance time on the app-name. The absence of significance between Groups A and B, and between Groups B and C, can be explained by the conditions that turned Group B into a hybrid Group. If the subjects decided to click on the privacy indicator, they were presented with the same explanation of the indicator that Group A had to read at the beginning of the study. However, if they (participants from Group B) refrain from doing so, they did not receive any further information about the privacy indicator, which is in line with the framework of Group C.

5.4 Indicator Evaluation by Participants

Overall, the privacy indicator was considered positive across all Groups, according to the collected feedback. The significant difference between Group A and C regarding the effect on the app-choice could be caused by the lack of explanation in Group C. While Group A was informed about the function and interpretation of the indicator by the preceding explanation of the privacy indicator in paper form, Group C had no opportunity to obtain more information about the indicator. From the video confrontation data, it can be concluded that this is the reason why some participants were skeptic about the indicator, its origin and its developer. Additionally, Group A might be exposed to suffer from low face validity, which should be taken into consideration. Based on the presented explanation of the privacy indicator, it is possible that the participants already suspected its' relevance to the experiment. On the other hand, Group B and C could have made a similar assumption only, because the privacy indicator was placed conspicuously on the app market.

5.5 Impact on Time to Make a Decision

On closer examination of the mean values regarding the decision times, it is noticeable that in the Video-call and Music scenarios, the participant required an average of just over 11 seconds to decide, while they needed considerably more time (approx. 28 seconds) to select an E-mail app. Another pair, in which similar decision times were found, is the scenario of the Weather and the News with just over 17 seconds each. All values from other pairs differ by almost two

seconds. Thus, it can be observed that the decision times between the scenarios differ significantly in most cases, which was confirmed in both the Friedman test and the Wilcoxon test. So, the participants in case of Music, Video-call and Messenger needed significantly less time to make a decision than in the scenarios concerning Games, Fitness and Email. As mentioned earlier, the familiarity of apps (or familiarity to the participant), has a high influence on the decision time as well as on selection behavior in general. The longest decision time in the E-mail scenario can be explained by the fact that the participants are mostly using the default E-mail app (from operating system provider), in this case 'Gmail' which was excluded in this study. The fact that important and official documents are sometimes sent in E-mail correspondence, could contribute to spending more time in making more appropriate selection. Comparing the differences in the decision times of the app in chronological order, significant effect between the difference in decision times of the first and second app can be found. However, this significant difference no longer exists after the Bonferroni correction. From the paired comparisons it is clear that only between Group B and D a significant effect in the difference in decision times between the first and second app could be found.

It can therefore, be assumed that the privacy indicator will not lead to a quicker decision even after several tasks have been performed with the indicator; rather it should lead an informed decision instead. However, it should be noted that the selection times of different scenarios (e.g. the Group Messenger and Weather), are combined into one variable, due to the fact that the participants carried out tasks in different sequences. Moreover, the subjects generally needed more time to complete the tasks in particular scenarios, such as the Weather scenario, than to complete the task in the Messenger scenario.

5.6 Further Impact on the Selection behavior

On closer examination of the selection of the app for fulfilling the Messenger task, it is noticeable that across all groups, most people have chosen an app that they already have installed and used on their own smartphone. The interviews showed that familiarity of the app had a major influence on selection behavior. Many of the participants justified their choice by pointing out that they already had the knowledge how to use the app and that the installation might have been easier to them.

Regarding the MUIPC, most of the participants were concerned about apps monitoring their mobile device and recording their activity on devices. In addition, they felt that their mobile device was sharing more privacy invasive information to the third-party than they wished. In summary, the participants were aware that their data is distributed through apps, which made them largely displeased.

Looking at the relationship between the choice of app and privacy concerns, it is noticeable that in most scenarios there are only a few correlations. However, in the E-mail scenario five positive correlations with a small effect can be seen. A positive correlation means that an increase in the score of the selected app is accompanied by a higher agreement with the statements of the MUIPC. So, a privacy concerned user is more likely to choose an app having less privacy invasive attribute.

6 LIMITATIONS AND FUTURE WORK

Here, we would like to discuss the limitations of our contributions and claims which are intended to be addressed in the future. This work is exploratory and it should be noted that the interpretation of null results in statistical tests also has limitation—a lack of detected statistical significance does not imply an absence of effect.

During the empirical study, we noticed that some apps were generally considered as safe by several participants, but marked as unsafe in our implemented app-store demo, e.g. Telegram due to vendors' advertisement on end to

end encryption. This partly confused the participants and perhaps had an impact on their decisions. So, there exists a probable error margin, because sometimes the app was selected despite facing the poor classification. Mostly this behavior is justified by the reputation and the private use of the app. However, a cross-correlation can be formulated between the users' perception from the MUIPC results and users' app selection behavior. We intend to counteract this limitation in near future by constructing a realistic indicator from app-behavior analysis data.

Another limitation was that the participants used glasses to operate the mobile phone. As a result, the eye movements in the video was recorded with poor quality. An attempt was made to counteract by asking the subject during the interview to describe more precisely on which characteristics was focused upon. So, we were compelled to exclude data sample from ten cases. However, we decided to limit the scope of this paper by excluding qualitative analysis of their opinion, and we intend to do it in future work.

This study did not recruit any individuals that were minors (aged below 18). Consequently, a large user base for Games, Social, Messaging and Lifestyle category remained out of consideration. Also, limited app selection options were offered to the participants, which could hinder in showcasing actual user behavior. We intend to conduct studies with unknown apps to avoid familiarity bias and with app ratings to determine comparative impact of the privacy indicator. Our plan for future work also includes, but are not limited to, implementation of multi-sourced metric for app-ranking mechanism, and investigations about *Privacy versus convenience*: is it more important for the user to have control over their data, or does convenience prevail? and *Intuitive contexts*: what are the life contexts and data types that can in meaningful ways get protected by ex-ante privacy indicator?

7 CONCLUSION

In this paper, we investigate the potential to introduce a simple privacy indicator within the ecosystem of smartphone apps. Our investigation includes designing and evaluating five variants of privacy indicators. Then we elaborate on the evaluation process of outperforming indicator and present results of a user study consisting of 82 participants which was focused on documenting their app selection behavior. The major findings from the collected data and statistical analysis of this study are: (i) An unilateral and simple privacy indicator is able to lead to a better judgment regarding the apps' trustworthiness and to provide ease of selection; (ii) Adequacy in privacy-preserving decision-making by subjects having indicator-illustrated interface, compared to the control group, can be observed; (iii) According to the collected feedback from the participants, adequacy of the indicator as a transparency and privacy enhancing tool can be confirmed. (iv) Results from the MUIPC analysis indicates that the participants expressed concerns, even though they acknowledge being aware about the nature of data collection by apps. (v) The eye-tracker data (recorded gaze-behavior video) shows that the indicator does not cause impact on the decision time. Hence, the equal time requirement (for both with and without indicator) can confirm that the user is able to make an informed decision with efficiency.

In addition, the result denotes that the indicator is generally perceived as positive and helpful. Feedback from participants point out that the indicator is able to initiate cautious thoughts and has a privacy-preserving influence on the selection behavior. However, it is also evident that app familiarity can cause bias in decision-making which can be inferred to the obvious factor—the marketplace facilitates a user to choose apps based on familiarity which can also be seen in the result from the questionnaire of the installed apps. Though privacy indicator poses requirement of a brief introduction for proper interpretation, it has the potential to bring greater use in avoiding privacy implications from apps' growing ability to collect, process and transmit data about the surroundings of the user.

REFERENCES

- [1] Alessandro Acquisti. 2009. Nudging privacy: The behavioral economics of personal information. *IEEE security & privacy* 7, 6 (2009), 82–85.
- [2] Alessandro Acquisti, Idris Adjerid, Rebecca Balebako, Laura Brandimarte, Lorrie Faith Cranor, Saranga Komanduri, Pedro Giovanni Leon, Norman Sadeh, Florian Schaub, Many Sleeper, Yang Wang, and Shomir Wilson. 2017. Nudges for Privacy and Security: Understanding and Assisting Users' Choices Online. *ACM Comput. Surv.* 50, 3, Article 44 (Aug. 2017), 41 pages. <https://doi.org/10.1145/3054926>
- [3] Alessandro Acquisti, Laura Brandimarte, and George Loewenstein. 2015. Privacy and human behavior in the age of information. *Science* 347, 6221 (2015), 509–514.
- [4] Manar Alohaly and Hassan Takabi. 2016. Better privacy indicators: a new approach to quantification of privacy policies. In *Twelfth Symposium on Usable Privacy and Security (SOUPS) 2016*.
- [5] Rebecca Balebako, Jaeyeon Jung, Wei Lu, Lorrie Faith Cranor, and Carolyn Nguyen. 2013. Little brothers watching you: Raising awareness of data leaks on smartphones. In *Proceedings of the Ninth Symposium on Usable Privacy and Security*. ACM, 12.
- [6] Izak Benbasat, Albert S Dexter, and Peter Todd. 1986. An experimental program investigating color-enhanced and graphical information presentation: an integration of the findings. *Commun. ACM* 29, 11 (1986), 1094–1105.
- [7] Richard E Christ. 1975. Review and analysis of color coding research for visual displays. *Human factors* 17, 6 (1975), 542–570.
- [8] Richard S Cimballo, Karen L Beck, and Donna S Sendziak. 1978. Emotionally toned pictures and color selection for children and college students. *The Journal of Genetic Psychology* 133, 2 (1978), 303–304.
- [9] European Commission. 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation). *Off J Eur Union* (2016), L119.
- [10] William J Conover and Ronald L Iman. 1981. Rank transformations as a bridge between parametric and nonparametric statistics. *The American Statistician* 35, 3 (1981), 124–129.
- [11] Android Developers Documentation. 2019. Android 9.0 changes. <https://developer.android.com/about/versions/pie/android-9.0-changes-all>. Accessed on 12-Sep-2019.
- [12] Android Developers Documentation. 2019. Dangerous permissions. https://developer.android.com/guide/topics/permissions/overview#dangerous_permission. Accessed on 12-Sep-2019.
- [13] Android Developers Documentation. 2019. Permissions overview. <https://developer.android.com/guide/topics/permissions/overview>. Accessed on 12-Sep-2019.
- [14] Android Developers Documentation. 2019. Runtime Permissions. <https://developer.android.com/distribute/best-practices/develop/runtime-permissions>. Accessed on 12-Sep-2019.
- [15] Adrienne Porter Felt, Elizabeth Ha, Serge Egelman, Ariel Haney, Erika Chin, and David Wagner. 2012. Android permissions: User attention, comprehension, and behavior. In *Proceedings of the eighth symposium on usable privacy and security*. ACM, 3.
- [16] Daniel Franzen and David Aspinall. 2016. PhoneWrap-Injecting the “How Often” into Mobile Apps. In *Proceedings of the 1st International Workshop on Innovations in Mobile Privacy and Security co-located with the International Symposium on Engineering Secure Software and Systems (ESSoS 2016)*. CEUR-WS.org, 11–19.
- [17] Lothar Fritsch and Nurul Momen. 2017. Derived Partial Identities Generated from App Permissions. In *Open Identity Summit (OID) 2017*. Gesellschaft für Informatik.
- [18] Ralph Gross and Alessandro Acquisti. 2005. Information revelation and privacy in online social networks. In *Proceedings of the 2005 ACM workshop on Privacy in the electronic society*. ACM, 71–80.
- [19] Jie Gu, Yunjie Calvin Xu, Heng Xu, Cheng Zhang, and Hong Ling. 2017. Privacy concerns for mobile app download: An elaboration likelihood model perspective. *Decision Support Systems* 94 (2017), 19–28.
- [20] Majid Hatamian, Nurul Momen, Lothar Fritsch, and Kai Rannenberg. 2019. A Multilateral Privacy Impact Analysis Method for Android Apps. In *Annual Privacy Forum*. Springer, 87–106.
- [21] Majid Hatamian, Jetzabel Serna, and Kai Rannenberg. 2019. Revealing the Unrevealed: Mining Smartphone Users Privacy Perception on App Markets. *Computers & Security* (2019). <https://doi.org/10.1016/j.cose.2019.02.010>
- [22] Majid Hatamian, Jetzabel Serna, Kai Rannenberg, and Bodo Igler. 2017. Fair: Fuzzy alarming index rule for privacy analysis in smartphone apps. In *International Conference on Trust and Privacy in Digital Business*. Springer, 3–18.
- [23] Judith H Hibbard and Ellen Peters. 2003. Supporting informed consumer health care decisions: data presentation approaches that facilitate the use of information in choice. *Annual review of public health* 24, 1 (2003), 413–433.
- [24] Jaeyeon Jung, Seungyeop Han, and David Wetherall. 2012. Short paper: enhancing mobile application permissions with runtime feedback and constraints. In *Proceedings of the second ACM workshop on Security and privacy in smartphones and mobile devices*. ACM, 45–50.
- [25] Patrick Gage Kelley, Joanna Bresee, Lorrie Faith Cranor, and Robert W. Reeder. 2009. A “Nutrition Label” for Privacy. In *Proceedings of the 5th Symposium on Usable Privacy and Security (Mountain View, California, USA) (SOUPS '09)*. ACM, New York, NY, USA, Article 4, 12 pages. <https://doi.org/10.1145/1572532.1572538>
- [26] Patrick Gage Kelley, Sunny Consolvo, Lorrie Faith Cranor, Jaeyeon Jung, Norman Sadeh, and David Wetherall. 2012. A conundrum of permissions: installing applications on an android smartphone. In *International conference on financial cryptography and data security*. Springer, 68–79.

- [27] Patrick Gage Kelley, Lorrie Faith Cranor, and Norman Sadeh. 2013. Privacy as part of the app decision-making process. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 3393–3402.
- [28] Lydia Kraus, Ina Wechsung, and Sebastian Möller. 2014. Using statistical information to communicate android permission risks to users. In *2014 Workshop on Socio-Technical Aspects in Security and Trust*. IEEE, 48–55.
- [29] William H Kruskal and W Allen Wallis. 1952. Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association* 47, 260 (1952), 583–621.
- [30] Hubert W Lilliefors. 1967. On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American statistical Association* 62, 318 (1967), 399–402.
- [31] Jialiu Lin, Bin Liu, Norman Sadeh, and Jason I. Hong. 2014. Modeling Users' Mobile App Privacy Preferences: Restoring Usability in a Sea of Permission Settings. In *10th Symposium On Usable Privacy and Security (SOUPS 2014)*. USENIX Association, Menlo Park, CA, 199–212. <https://www.usenix.org/conference/soups2014/proceedings/presentation/lin>
- [32] Bin Liu, Mads Schaarup Andersen, Florian Schaub, Hazim Almuhiemedi, Shikun (Aerin) Zhang, Norman Sadeh, Yuvraj Agarwal, and Alessandro Acquisti. 2016. Follow My Recommendations: A Personalized Privacy Assistant for Mobile App Permissions. In *Twelfth Symposium on Usable Privacy and Security (SOUPS 2016)*. USENIX Association, Denver, CO, 27–41. <https://www.usenix.org/conference/soups2016/technical-sessions/presentation/liu>
- [33] Patrick E McKnight and Julius Najab. 2010. Mann-Whitney U Test. *The Corsini encyclopedia of psychology* (2010), 1–1.
- [34] George A Miller. 1956. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review* 63, 2 (1956), 81.
- [35] Nurul Momen. 2018. Towards Measuring Apps' Privacy-Friendliness. *Licentiate Dissertation, Karlstad University* (2018).
- [36] Nurul Momen, Sven Bock, and Lothar Fritsch. 2020. Accept-Maybe-Decline: Introducing Partial Consent for the Permission-based Access Control Model of Android. In *Proceedings of the 25th ACM Symposium on Access Control Models and Technologies*. 71–80.
- [37] Nurul Momen and Lothar Fritsch. 2020. App-generated digital identities extracted through Android permission-based data access—a survey of app privacy. *SICHERHEIT 2020* (2020).
- [38] Nurul Momen, Majid Hatamian, and Lothar Fritsch. 2019. Did App Privacy Improve After the GDPR? *IEEE Security & Privacy* 17, 6 (2019), 10–20.
- [39] Nurul Momen, Tobias Pulls, Lothar Fritsch, and Stefan Lindskog. 2017. How much Privilege does an App Need? Investigating Resource Usage of Android Apps. In *The Fifteenth International Conference on Privacy, Security and Trust—PST 2017. August 28-30, 2017 Calgary, Alberta, Canada*. IEEE.
- [40] Patrick Murrmann and Simone Fischer-Hübner. 2017. Tools for achieving usable ex post transparency: a survey. *IEEE Access* 5 (2017), 22965–22991.
- [41] Marie-Therese Puth, Markus Neuhäuser, and Graeme D Ruxton. 2015. Effective use of Spearman's and Kendall's correlation coefficients for association between two measured traits. *Animal Behaviour* 102 (2015), 77–84.
- [42] Prashanth Rajivan and Jean Camp. 2016. Influence of Privacy Attitude and Privacy Cue Framing on Android App Choices. In *Twelfth Symposium on Usable Privacy and Security (SOUPS 2016)*. USENIX Association, Denver, CO. <https://www.usenix.org/conference/soups2016/workshop-program/wpi/presentation/rajivan>
- [43] Kopo M Ramokapane, Anthony C Mazeli, and Awais Rashid. 2019. Skip, Skip, Skip, Accept!!!: A Study on the Usability of Smartphone Manufacturer Provided Default Features and User Privacy. *Proceedings on Privacy Enhancing Technologies* 2019, 2 (2019), 209–227.
- [44] Sanae Rosen, Zhiyun Qian, and Z Morely Mao. 2013. Appprofiler: a flexible method of exposing privacy-related behavior in android applications to end users. In *Proceedings of the third ACM conference on Data and application security and privacy*. ACM, 221–232.
- [45] Stefan Schneegass, Romina Poguntke, and Tonja Machulla. 2019. Understanding the Impact of Information Representation on Willingness to Share Information. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (Glasgow, Scotland Uk) (CHI '19)*. ACM, New York, NY, USA, Article 523, 6 pages. <https://doi.org/10.1145/3290605.3300753>
- [46] Fuming Shih, Ilaria Lippardi, and Daniel Weitzner. 2015. Privacy Tipping Points in Smartphones Privacy Preferences. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (Seoul, Republic of Korea) (CHI '15)*. ACM, New York, NY, USA, 807–816. <https://doi.org/10.1145/2702123.2702404>
- [47] Irina Shklovski, Scott D Mainwaring, Halla Hrund Skúladóttir, and Höskuldur Borgthorsson. 2014. Leakiness and creepiness in app space: Perceptions of privacy and mobile app use. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*. ACM, 2347–2356.
- [48] Christopher Thompson, Maritza Johnson, Serge Egelman, David Wagner, and Jennifer King. 2013. When it's better to ask forgiveness than get permission: attribution mechanisms for smartphone resources. In *Proceedings of the Ninth Symposium on Usable Privacy and Security*. ACM, 1.
- [49] Jason Watson, Heather Richter Lipford, and Andrew Besmer. 2015. Mapping user preference to privacy default settings. *ACM Transactions on Computer-Human Interaction (TOCHI)* 22, 6 (2015), 32.
- [50] Primal Wijesekera, Arjun Baokar, Ashkan Hosseini, Serge Egelman, David Wagner, and Konstantin Beznosov. 2015. Android permissions remystified: A field study on contextual integrity. In *24th {USENIX} Security Symposium ({USENIX} Security 15)*. 499–514.
- [51] Heng Xu, Sumeet Gupta, Mary Beth Rosson, and John M Carroll. 2012. Measuring mobile users' concerns for information privacy. (2012).