HAL open science

Bayesian Evaluation of User App Choices in the Presence of Risk Communication on Android Devices

B. Momenzadeh, S. Gopavaram, S. Das, L. J. Camp

To cite this version:

B. Momenzadeh, S. Gopavaram, S. Das, L. J. Camp. Bayesian Evaluation of User App Choices in the Presence of Risk Communication on Android Devices. 14th International Symposium on Human Aspects of Information Security and Assurance (HAISA), Jul 2020, Mytilene, Lesbos, Greece. pp.211-223, 10.1007/978-3-030-57404-8_16 . hal-03657733

HAL Id: hal-03657733

https://inria.hal.science/hal-03657733v1

Submitted on 3 May 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

Bayesian Evaluation of User App Choices in the Presence of Risk Communication on Android Devices

B. Momenzadeh[1], *S.Gopavaram*[1], *S.Das*[1,2], *and L.JCamp*[1]

1. Indiana University Bloomington
2. University of Denver
(smomenza, sgopavar, sancdas, ljcamp)@iu.edu

Abstract

In the age of ubiquitous technologies, security- and privacy-focused choices have turned out to be a significant concern for individuals and organizations. Risks of such pervasive technologies are extensive and often misaligned with user risk perception, thus failing to help users in taking privacy-aware decisions. Researchers usually try to find solutions for coherently extending trust into our often inscrutable electronic networked environment. To enable security- and privacy-focused decision-making, we mainly focused on the realm of the mobile marketplace, examining how risk indicators can help people choose more secure and privacy-preserving apps. We performed a naturalistic experiment with $N = 60$ participants, where we asked them to select applications on Android tablets with accurate real-time marketplace data. We found that, in aggregate, app selections changed to be more risk-averse in the presence of user risk-perception-aligned visual indicators. Our study design and research propose practical and usable interactions that enable more informed, risk-aware comparisons for individuals during app selections. We include an explicit argument for the role of human decision-making during app selection, beyond the current trend of using machine learning to automate privacy preferences after selection during run-time.

Keywords

Mobile App Permissions, Android, Risk Communication, Human-Centered Privacy and Security, Mobile Security.

1. Introduction

Permissions models are an excellent initiative to inform smartphone users of the services that each application might access. However, research has shown that they have failed to consistently communicate the privacy and security risks of apps on mobile platforms [1, 4, 8]. Currently, many researchers are discarding permissions as futile user communication, focusing on implicit instead of explicit choices and using machine learning or agent-based permissions management after installation [13, 17]. Not only does much research in this area focus on building machine learning tools that regulate resources accessed by apps during runtime, but Android OS has also shifted from app permission manifests to runtime permissions. Mitigating privacy risks for apps during runtime is essential, and much of this mitigation must be automated. However, an automated system during runtime has its own limitations.

Our work motivation is to determine if a multi-level communication system can support explicit individual decision-making during app selection. In addition to supporting individual autonomy, privacy-aware decision making at the time of application election offers promise for the entire ecosystem. Supporting individual risk-aware decisions in app selection could enable app providers to differentiate themselves in the app marketplace and provide developers with an incentive to consider user privacy when building apps. In this paper, we focus on enhancing the decision-time communication of risks to the user. We

built a risk-indicative warning system and tested it with an operational app store in a natural environment. This warning system is built upon the findings of previous research in usable security on mobile devices and behavioral psychology.

The essential contribution of our work is an empirical illustration of the changes in participants' decision-making when provided with simple, timely, comprehensible warnings. Instead of removing permissions interactions, an alternative approach is to improve the communication and design aspect to enable users to take privacy-aware and security-aware decisions. Specifically, we illustrate the efficacy of a multi-level system where information is immediately available and summarized, with the option of searching for additional information. As this is a recommendation for design in general practice and for warning systems specifically, this is not surprising [18]. In addition we use a Bayesian experiment design and analysis to compare the distribution of app selections from participants in our market to those in the standard app market. The purpose of using a Bayesian approach is to test the interaction in a noisy, confounded naturalistic environment and to provide a stronger confidence measure than a traditional means comparison.

2. Background Motivation

2.1. Permission Models on Mobile Phones

To develop our warning system, we leveraged the Android permissions model that was used before Android OS moved to the runtime model (which was previously only used by iOS). Instead of presenting the list of permissions immediately, we added a layer of interaction that summarizes the risk of the agreed-upon permissions by the users.

Empirical research has found a significant lack of understanding, not only about the implications of providing sensitive permissions but also about the underlying meaning of permissions [4, 8]. Smartphone users are mostly unaware of the resources accessed by apps [12]. For permission manifests used in Android, repeated research has shown that people usually ignore or pay little attention to them; for example, a series of online surveys and laboratory studies conducted by Felt et al. found that only 17% of the participants paid attention to permissions during app installation [4]. Another study conducted by Rajivan et al. four years later found that only 13% of the participants viewed the permissions by clicking on them [14].

In recognition of the fact that previous permissions models were inadequate, there has been a move to automate permissions decisions based on machine learning models of observed user behavior. Models of user preferences may be driven by background observations, possibly augmented by explicit queries about acceptable data use [13, 17]. The addition of machine learning mitigates risk, but it does not enable purposeful choice. Those who value their privacy are unable to make privacy-preserving app selections, as there is a lack of adequate decision-making support at the moment of selection [2].

2.2. Visual Indicators

We based the design of our visual warning risk-indicator for aggregate privacy on previous work and chose padlocks. We decided to frame the indicator positively, so more padlocks implied a lower risk. For that reason, we refer to these ratings as *privacy ratings*. We considered the five principles proposed by Rajivan et al. [14]. First we selected *icons aligned with user mental models of security*, meaning we selected the widely-used lock icon from HTTPS. Second, given that *privacy communicating icons should be in terms of*

privacy offered by the app/software, we based ratings on the permissions. Third, we made the *scale of privacy communicating icons consistent with other indicators*. Fourth, and this inherently aligns with our design, *icons should be presented early in the decision-making process, while people compare apps to choose and install*. And the fifth principle, that *privacy communication should be trustworthy*, was embedded in our use of permissions for rating the apps.

3. Methodology

The goal of the experiment is to investigate if the introduction of proposed visual indicators in an actual PlayStore would change user app selections. Thus, we built an alternate PlayStore and asked participants to select multiple apps from different categories. We then ranked the apps presented to the users based on the number of downloads they received in the experiment. We compared these rankings against the download-based rankings in the actual PlayStore. Through this study design, we aimed at answering the following research questions (RQ):

RQ1 In the absence of differing privacy options, do our participants make choices that are indistinguishable from the Android Marketplace?

RQ2 When the functionality of the apps is the same, but the privacy options differ, do our participants make choices that are indistinguishable from the Android Marketplace?

RQ3 When both functionality and privacy options vary, do our participants make choices that are indistinguishable from the Android Marketplace?

3.1. Alternate PlayStore

To answer the research questions mentioned above, we built a functional app store with real-world applications, app ratings, and download counts. The user interface for our app store resembled that of Google's PlayStore on Android Jelly Bean (Version 4.1). Unlike Google's PlayStore, our app store presented users with visual indicators for aggregate risk (privacy ratings). We derived the privacy ratings from PrivacyGrade [10, 11] [1]. PrivacyGrade generated privacy grades ranging from A through D for apps on Android. We retrieved the privacy grade and converted it into a numerical rating between 1 and 5. Since we use positive framing, a privacy rating of 5 is equivalent to an A grade. The privacy rating is presented on both the *list of apps* page and the *app description* page. Figures 1 and 2 show the *list of apps* and the *app description* pages respectively alongside their counterparts from the actual PlayStore. We added a button at the top of the description page which would show permissions to the participants. We used this button to track which participants viewed the permissions.

We built the alternative PlayStore by modifying the code of *BlankScore* [2] (An open-source Google PlayStore client) and used an open-source API to query Google's servers for information. The alternative PlayStore enabled us to provide accurate user ratings, download counts, descriptions, and a list of permissions of apps. Additionally, search results for applications on the alternative PlayStore were the same as the results on the ~~actual PlayStore, including the order~~ of presentation of apps.

[1] http://privacygrade.org/

[2] https://github.com/mar-v-in/BlankStore

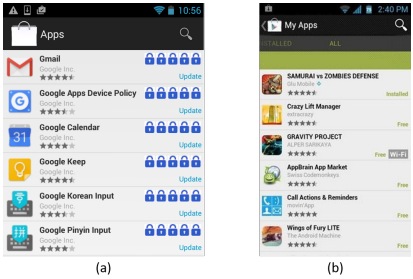

Figure 1: (a) *List of apps* page on the alternate PlayStore with privacy ratings on the alternate PlayStore. (b) *List of apps* page with no risk score on the actual PlayStore

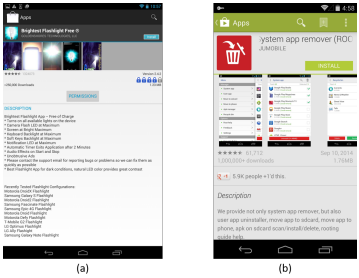

Figure 2: (a) *App description* page on the alternate PlayStore with privacy ratings on the alternate PlayStore. (b) *App description* page with no risk score on the actual PlayStore

3.2. Study Design

We recruited a total of $N = 60$ participants for the experiment through our outreach at the public library and the local farmers market to obtain socio-economically and culturally homogeneous population. A core design goal was to make the experimental interaction as close as possible to the experience of interacting with the Android PlayStore. We installed the alternate PlayStore on Nexus 7 tablets and provided those tablets to our participants. We then provided each of our participants with a list of keywords to search for on the alternate PlayStore. These keywords correspond to the app categories we chose for our experiment. Each search provided a list of up to 16 apps for a given category, and we asked the participants to select and download 4 of them. To make sure all the participants saw the same results, we ensured that each participant used the same category names. The search results for all the keywords on the alternate PlayStore were identical to the ones generated by the actual PlayStore. We did not describe the purpose of the experiment, mention security, nor describe the indicators to the participants beforehand.

3.3. Statistical Analysis Approach

We used a clinical research model with an observational study by selecting a subset of participants and exposing them to an experimental condition then comparing their outcomes with the large known set of results without the condition [15]. We compared the means of the two groups, using a posthoc Tukey pairwise comparison. We include these results for each category for the ease of comparison with other work; however, we also argue that the lack of nuance in means comparisons argues for the use of a Bayesian approach. The Kruskal-Wallis Test shows the significance of differences in weighted means of privacy ratings for the four categories, which are: 0.005 for Games; 0.53 for Flashlights; 0.02 for Photos; and 0.28 for Weather. The results of this comparison show the significance of the differences between the mean privacy rating of apps chosen by those using our experimental PlayStore and the mean privacy rating of apps chosen through the actual Android PlayStore. We calculated a region of practical equivalence (ROPE) based on a Highest Density Interval (HDI) of 95%, which means the area that contains the 95% most credible values for participants' choices have the same distribution of the users' choices in the PlayStore. The comparison is between the behavior of our experimental sample and the behavior of people using the regular Android PlayStore. The advantages of an analysis using a Bayesian approach are that it integrates historical information and that it is valid

with a small sample size. Other advantages are that the Bayesian analysis requires no assumptions about normality or distribution of the data. When examining the difference of means graphs we provide in reporting our Bayesian analysis, the dotted line on zero marks the point where the distributions match. The black line underneath the bars defines the 95% HDI area. The numbers on either side of the black line to specify the start and end threshold of the 95% interval. The bars show the distribution of the Difference of Means data. We will talk about each graph individually below as we report the results for each of the four categories.

3.4. Results

Out of the 60 participants, 58% were male, and 42% were female. After completing the experiment, all the participants in the study were asked to answer questions related to their app installation behavior. One of the questions asked them about the criteria they considered when selecting an application. In response to this question, 48% of the participants stated that they prioritized an app's features over other criteria when selecting an app to install. After that, the popularity ranked second, and friends' suggestion was the third choice. The other criteria, in this case, were ads, permissions, rank, reviews, and design. The survey inquired about permissions behaviors, asking participants about how often they checked an app's permissions. To this, 22% of the participants stated that they check permissions "almost every time" or "always" when installing an app. We also asked participants if they had previously refused to continue with the installation of an app because of its permissions. 79.6% of the participants stated that they had refused to install an app because of the permissions it requested in their real life. However, in practice, there was only one instance where a participant did not continue with the installation of an app because of its permissions (or after viewing the permissions). Only 7% of the total installations preceded with a check of the app's permissions in our experiment. This discrepancy between the observed and stated behavior is consistent with previous research studies [14, 19].

We also investigated if the addition of aggregate risk information was cognitively burdensome for our participants. Therefore, we used the NASA TLX instrument to measure the mental workload involved in using the alternate PlayStore to select apps [6]. The results indicate that the majority of the participants (78%) found the workload to be minimal.

3.4.1. Research Question 1: Are Participants Representative?

We chose the Flashlight category to address this research question. It is not that flashlights are particularly safe and secure; rather, such apps all have the same level of privacy in terms of permissions. Table 1 shows both participant selections, PlayStore selections, and the similar privacy ratings of all of the apps. All flashlight apps also have the same functionality. When there is no difference in the privacy indicators nor the functionality of the apps, we wanted to explore whether the selections of our participants were different from those in the PlayStore. Using Bayesian analysis we show that our participants were indistinguishable from a random sample of people selecting Android apps.

To address the frequentist results first, the weighted average privacy ratings of the flashlight apps in the PlayStore is 4.94. The weighted app ratings were 4.52 and 4.51 for the store and the experimental participants, respectively. In Figure 3, 0 is marked with a dotted line. Zero falls near the center of the region of practical equivalence. The entire HDI is within the ROPE, so the difference is practically equivalent to the null value. In the case of the Flashlight apps, our participants were statistically indistinguishable from a random

App Name	Downloads	Locks	Exp. Rank	PlayStore Rank
Super-Bright LED Flashlight	38	5	1	1
Color Flashlight	34	5	2	3
Tiny Flashlight + LED	26	5	3	2
Brightest Flashlight Free	20	4	4	4
Flashlight Galaxy S7	16	5	5	10
Flashlight Galaxy	16	5	5	9
Brightest LED Flashlight	15	5	7	5
Flashlight	12	5	8	11
High-powered Flashlight	11	5	9	6
Flashlight Widget	7	5	10	12
FlashLight	6	5	11	7
Flashlight for HTC	5	5	12	13
Flashlight	3	5	13	8

Table 1: Flashlight Category by order of Downloads in the Experiment: Apps' Rank in the PlayStore, Downloads in the Experiment, and Privacy Rating (Locks)

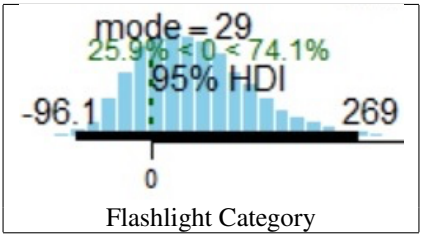

Figure 3: Regions of Practical Equivalence for the selection of flashlight apps (RQ1). The comparison of the selections in our sample (μ_1) and the selections of our participants (μ_2). The graph shows the ROPE for the difference of means ($\mu_1 - \mu_2$), showing that our participants' selections were indistinguishable from a random sample of Android app purchasers.

sample of the selections made in the larger PlayStore. This verifies that, in the absence of differing privacy ratings, our participants' choices were indistinguishable from those of a random sample of Android users.

3.4.2. Research Question 2: Similar Functionality and Different Privacy Rating

To answer RQ2, we used the apps in the Weather category. In many cases, we expected that the individuals would trade privacy or security for some other feature-based benefit. Given the difficulty in measuring how individuals value risk avoidance, we sought a category with little functional variance and high variability in information risk. To begin with an illustrative frequentist comparison, the weighted average privacy ratings of the weather apps in the PlayStore were 4.26 and 4.25 for our participants. The weighted average app ratings were 4.39 for both PlayStore users and experimental participants. The Kruskal-Wallis difference in means had a p-value of 0.28. (Note that Kruskal-Wallis examines the contrast of the ways, while a Bayesian approach considers the likelihood of a distribution). The dominance of the most popular weather app, with a privacy rating of four, results in a slight skewing of the results. In many contexts, it is well-understood that people select the first choice on a list or go with defaults [7, 9]. The overall difference in the means between weather apps is shown in Figure 4. This shows little overlap between the distribution of selected weather apps between our participants and the distribution of participants in the PlayStore. That is, the likelihood that the selection of apps by our participants is an unbiased distribution resulting from a sample of the prior distribution, as shown by the PlayStore, is practically equivalent to the null value. This result indicates that the distribution is biased, and thus our warning visual risk-indicator affected our participants' choices.

App Name	Downloads	Locks	Exp. Rank	PlayStore Rank
Weather - The Weather Channel	40	4	1	1
AccuWeather	31	5	2	2
Yahoo Weather	27	5	3	5
MyRadar Weather Radar	27	5	3	10
Weather Underground	19	5	5	11
Weather by WeatherBug	16	3	6	6
Weather & Clock Widget Android	14	4	7	4
Transparent clock & weather	11	3	8	6
NOAA Weather Unofficial	7	4	9	12
Go Weather Forecast	5	4	10	3
Weather Project	5	1	10	15
Weather, Widget Forecast Radar	3	4	12	8
Weather Project	2	1	13	14
iWeather-The Weather Today	2	1	13	13
Weather	1	4	15	9

Table 2: Weather Category by order of Downloads in the Experiment: Apps' Rank in the PlayStore, Downloads in the Experiment, and Privacy Rating (Locks)

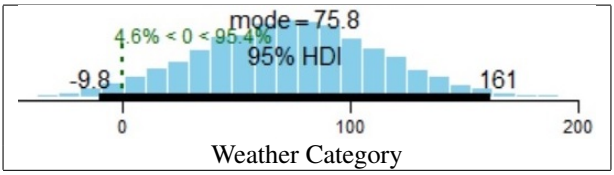

Figure 4: Regions of Practical Equivalence, showing the difference between our participants (μ_1) and the selections in the Android marketplace (μ_2). The graph shows the ROPE for the difference of means ($\mu_1 - \mu_2$), showing less than 5% overlap with 0 region and an unskewed distribution.

3.4.3. Research Question 3: Varying Functionality and Privacy Rating

Our other two categories, Photos and Games, were used to answer this research question. In photos, there is more variance in functionality than in weather or flashlight apps. Photo apps coordinate with different services (e.g., Instagram or Facebook), offer different filters (e.g., glitter, party hats, sepia tones), different functionality (e.g., annotating), and different sharing modes. The weighted average privacy ratings of the game apps in the PlayStore and of the choices of participants were both 4.93. The weighted app ratings were 4.43 and 4.34 for the store and the experimental participants, respectively. The Bayesian analysis of the Games category is shown in Figure 5. The dotted line falls into the 95% HDI. The volume of the ROPE that intersects with the likelihood of this being the practical equivalent of a random, unbiased sample of the prior known PlayStore distribution is 4.6%, approaching 5%. The initial value of the HDI (-58.5) is comparable to the distance between the zero and mode, which falls on 102. The average privacy rating of the photos apps in the PlayStore is 4.69. The weighted average privacy ratings of the choices of participants were 4.97. The weighted app ratings were 4.39 and 4.41.

In the case of photo apps, the distribution of app ratings and risk was such that individuals could mitigate risk without sacrificing any benefits. With photo applications, participants chose more secure apps over other more popular apps with more downloads, more familiarity, and more popular design. *PicsArt Photo Studio and Collage* was particularly selected by only three of our participants in our experiment for the photos category, while it was the second-ranked app in terms of the number of downloads with this search term in Google PlayStore. The results for the Photos category is quite similar to the Weather category. Zero falls at the beginning of the HDI interval. However, the distance to the start of the range is insignificant, and the distance to mode is also significant. This implies that we have influenced participants' decisions in this category as well and that participants'

App Name	Downloads	Locks	Exp. Rank	PlayStore Rank
Fruit Ninja Free	39	5	1	2
Subway Surfers	23	5	2	1
Super Smash Jungle World	22	5	3	8
PAC-MAN	20	5	4	5
Wheel of Fortune Free Play	16	5	5	13
Color Switch	15	5	6	7
Piano Tiles 2™	15	5	6	4
slither.io	12	5	8	3
Rolling Sky	11	5	9	6
Block! Hexa Puzzle	4	1	10	9
Flip Diving	3	5	11	10
Battleships - Fleet Battle	2	5	12	17
Snakes & Ladders King	2	5	12	11
Board Games	1	5	14	13
Best Board Games	1	5	14	15
Checkers	1	5	14	12
Mancala	1	3	14	16

Table 3: Games Category by order of Downloads in the Experiment: Apps' Rank in the PlayStore, Downloads in the Experiment, and Privacy Rating (Locks)

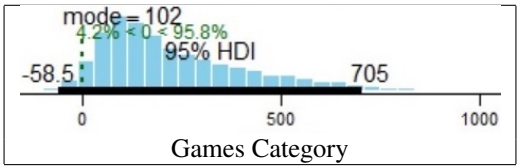

Figure 5: ROPE for the difference of means ($\mu_1 - \mu_2$) for our participants and the Android marketplace in selection of game apps. Note the distribution of probabilities is highly skewed, decreasing confidence, while the overlap is 4.2% (RQ3)

choices could be distinguished from options in the PlayStore. The results show that the likelihood that the parameters that characterize the distribution of the choices by our participants cannot reasonably be considered the same as the parameters that characterize the distribution of apps chosen by those in the PlayStore.

App Name	Downloads	Locks	Exp. Rank	PlayStore Rank
Google Photos	39	5	1	1
PhotoDirector Photo Editor App	25	5	2	9
Photo Lab Picture Editor FX	24	5	3	5
Gallery	23	5	4	10
Photo Editor Pro	20	5	5	4
A+ Gallery Photos	19	5	6	12
Photo Collage Editor	17	5	7	5
PhotoGrid & Photo Collage	15	5	8	3
Toolwiz Photos - Pro Editor	13	5	9	11
Photo Editor Collage Maker Pro	9	5	10	7
PicsArt Photo Studio	3	3	11	2
Phonto - Text on Photos	1	5	12	8

Table 4: Photos Category by order of Downloads in the Experiment: Apps' Rank in the PlayStore, Downloads in the Experiment, and Privacy Rating (Locks)

3.5. Discussion

Can we use the most common indicator of privacy on the Internet– a lock– to communicate aggregate information about mobile app privacy risk? If so, would this change the choices made by participants in the marketplace? There is no a priori answer. Information about privacy and security risks could be ignored or unwelcome. Studies in risk communication have shown that individuals find risk more acceptable if the exposure to the risk is voluntary, and when the individual exposed is capable of avoiding the risk or freely choosing it [5]. That is, shifting the nexus of control may actually increase aggregate risk-taking;

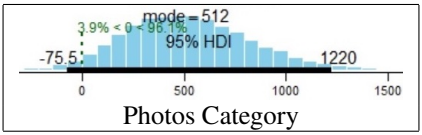

Figure 6: Regions of Practical Equivalence for the difference of means ($\mu_1 - \mu_2$) for our participants and the Android marketplace in selection of photo apps (RQ3)

the perception of control increases data sharing [16]. In privacy, this response is called the 'control dilemma' [3]. To address these questions in the context of mobile apps, we asked the research questions described above.

First (RQ1), can we confirm that our group of participants are indistinguishable from the Google's PlayStore users as a whole when presented with apps that had the same kind of functionality and the same privacy ratings? We used the Flashlight category for this purpose. We found that the selection of apps under this condition was indistinguishable from a random sample of app selections in the PlayStore. Our next question (RQ2) is if the participants would make different choices compared to that of Google's PlayStore users in the presence of variable ratings given the same functionality. For this question, we used the Weather category. The results from this category showed that choices in our experiment are significantly different from those made in the Google PlayStore, thus offering a high level of confidence that the ratings influenced user decision-making. Finally, we ask (in RQ3) if we can be confident that the participants' decisions are different from Google's PlayStore users when the functionality and the privacy ratings both vary. We used Games and Photos as the categories to answer this question. The Photos category indicates that the inclusion of privacy ratings, even with marginal rating differences, results in a different distribution of apps selected. In Games, we have a low level of confidence that the participants' decisions are different and can not conclude the parameters are changed.

In summary, we built a functional app store with real, accurately rated apps and added visual indicators for aggregate risk using the padlock icon. The functional app store simulation made it possible for us to compare the choices of the participants directly with those of people using Google's PlayStore. It is true that using a functional app store with real-world applications meant that we were not able to adequately control for biases, including ordering, familiarity, and reputation. However, we are confident that our participants' choices when there was no variance in privacy is indistinguishable from those of the PlayStore at large, providing confidence in the representativeness of our sample that is difficult to obtain in a traditional controlled laboratory experiment.

4. Acknowledgement

This research was supported in part by the National Science Foundation under CNS 1565375, Cisco Research Support, and the Comcast Innovation Fund. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s). They do not necessarily reflect the views of the U.S. Government, NSF, Cisco, Comcast, Indiana U, or the University of Denver.

5. References

[1] Y. Agarwal and M. Hall. ProtectMyPrivacy: Detecting and Mitigating Privacy Leaks on iOS Devices Using Crowdsourcing. In *Proceeding of the 11th Annual International Conference on Mobile Systems, Applications, and Services*, pages 97–110. ACM, 2013.

[2] R. Böhme, S. Koble, and T. Dresden. On the Viability of Privacy-Enhancing Technologies in a Self-Regulated Business-to-Consumer Market: Will Privacy Remain a Luxury Good? In *6th Workshop on the Economics of Information Security (WEIS)*, 2007.

[3] L. Brandimarte, A. Acquisti, and G. Loewenstein. Misplaced Confidences Privacy and the Control Paradox. *Social Psychological and Personality Science*, 4(3):340–347, 2013.

[4] A. P. Felt, E. Ha, S. Egelman, A. Haney, E. Chin, and D. Wagner. Android permissions: user attention, comprehension, and behavior. In *8th Symposium on Usable Privacy and Security (SOUPS)*, pages 3:1–3:14. ACM, 2012.

[5] V. Garg and J. Camp. End User Perception of Online Risk Under Uncertainty. In *45th Hawaii International Conference on System Science (HICSS)*, pages 3278–3287. IEEE, 2012.

[6] S. G. Hart and L. E. Staveland. Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. *Advances in Psychology*, 52:139–183, 1988.

[7] E. J. Johnson, S. Bellman, and G. L. Lohse. Defaults, Framing and Privacy: Why Opting In-Opting Out. *Marketing Letters*, 13(1):5–15, 2002.

[8] P. G. Kelley, S. Consolvo, L. F. Cranor, J. Jung, N. Sadeh, and D. Wetherall. A Conundrum of Permissions: Installing Applications on an Android Smartphone. In *16th International Conference on Financial Cryptography and Data Security*, pages 68–79. Springer, 2012.

[9] Y.-L. Lai and K.-L. Hui. Internet Opt-In and Opt-Out: Investigating the Roles of Frames, Defaults and Privacy Concerns. In *14th ACM SIGMIS CPR Conference on Computer Personnel Research: Forty Four Years of Computer Personnel Research: Achievements, Challenges & the Future*, pages 253–263. ACM, 2006.

[10] J. Lin, S. Amini, J. I. Hong, N. Sadeh, J. Lindqvist, and J. Zhang. Expectation and Purpose: Understanding Users' Mental Models of Mobile App Privacy through Crowdsourcing. In *14th ACM Conference on Ubiquitous Computing*, pages 501–510. ACM, 2012.

[11] J. Lin, B. Liu, N. Sadeh, and J. I. Hong. Modeling Users' Mobile App Privacy Preferences: Restoring Usability in a Sea of Permission Settings. In *10th Symposium on Usable Privacy and Security (SOUPS)*, pages 199–212. ACM, 2014.

[12] A. Mylonas, A. Kastania, and D. Gritzalis. Delegate the Smartphone User? Security Awareness in Smartphone Platforms. *Computers & Security*, 34:47 – 66, 2013.

[13] K. Olejnik, I. Dacosta, J. S. Machado, K. Huguenin, M. E. Khan, and J.-P. Hubaux. Smarper: Context-Aware and Automatic Runtime-Permissions for Mobile Devices. In *38th IEEE Symposium on Security and Privacy (SP)*, pages 1058–1076. IEEE, 2017.

[14] P. Rajivan and J. Camp. Influence of Privacy Attitude and Privacy Cue Framing on Android App Choices. In *12th Symposium on Usable Privacy and Security (SOUPS)*. USENIX, 2016.

[15] B. Röhrig, J.-B. Du Prel, D. Wachtlin, and M. Blettner. Types of Study in Medical Research: Part 3 of a Series on Evaluation of Scientific Publications. *Deutsches Ärzteblatt International*, 106(15):262, 2009.

[16] F. Stutzman, R. Gross, and A. Acquisti. Silent Listeners: The Evolution of Privacy and Disclosure on Facebook. *Journal of Privacy and Confidentiality*, 4(2):2, 2013.

[17] P. Wijesekera, A. Baokar, L. Tsai, J. Reardon, S. Egelman, D. Wagner, and K. Beznosov. The Feasibility of Dynamically Granted Permissions: Aligning Mobile Privacy with User Preferences. In *38th IEEE Symposium on Security and Privacy (SP)*, pages 1077–1093. IEEE, 2017.

[18] M. S. Wogalter, D. DeJoy, and K. R. Laughery. *Warnings and risk communication*. CRC Press, 1999.

[19] L. Yang, N. Boushehrinejadmoradi, P. Roy, V. Ganapathy, and L. Iftode. Short Paper: Enhancing Users' Comprehension of Android Permissions. In *2nd ACM Workshop on Security and Privacy in Smartphones and Mobile Devices*, pages 21–26. ACM, 2012.