Evaluating Users' Perceptions about a System's Privacy: Differentiating Social and Institutional Aspects

Oshrat Ayalon and Eran Toch, ***Tel Aviv University***

https://www.usenix.org/conference/soups2019/presentation/ayalon

This paper is included in the Proceedings of the Fifteenth Symposium on Usable Privacy and Security.

August 12–13, 2019 • Santa Clara, CA, USA

ISBN 978-1-939133-05-2

Open access to the Proceedings of the Fifteenth Symposium on Usable Privacy and Security is sponsored by USENIX.

Evaluating Users' Perceptions about a System's Privacy: Differentiating Social and Institutional Aspects

Oshrat Ayalon, *Tel Aviv University* Eran Toch, *Tel Aviv University*

Abstract

System design has a crucial effect on users' privacy, but privacy-by-design processes in organizations rarely involve end-users. To bridge this gap, we investigate how User-Centered Design (UCD) concepts can be used to test how users perceive their privacy in system designs. We describe a series of three online experiments, with 1,313 participants overall, in which we attempt to develop and validate the reliability of a scale for Users' Perceived Systems' Privacy (UPSP). We found that users' privacy perceptions of information systems consist of three distinctive aspects: institutional, social and risk. We combined our scale with A/B testing methodology to compare different privacy design variants for given background scenarios. Our results show that the methodology and the scale are mostly applicable for evaluating the social aspects of privacy designs.

1. Introduction

System designs that do not meet the users' privacy expectations can startle users and lead them to abandon the system altogether [16, 20, 41, 50]. For example, in Felt et al. study, a participant reported about uninstalling an app after it had used his/her contact list information to send spam texts and emails [20]. These examples of mis-design highlight the importance of designing systems with privacy from the ground up, as promised by the Privacy-by-Design (PbD) approach. It calls for implementing privacy mechanisms in the systems at the initial stages of the development process to create privacy-respectful systems in advance [13, 38]. While PbD is part of official guidelines by the FTC and by the recent European General Data Protection Regulation (GDPR) [21], it is also criticized of being too focused on compliance to privacy regulation, rather than on providing the best privacy design to the users [66]. As a response, Koops et al. argue for broadening the envelope of PbD, fostering "the right set of mindset of those responsible for developing and running data processing systems." [34]

USENIX Symposium on Usable Privacy and Security (SOUPS) 2019. August 11 -- 13, 2019, Santa Clara, CA, USA.

End-users' long-term concerns and expectations are not always considered in the process of designing the privacy characteristics and features in systems. Therefore, we argue that privacy-by-design processes should take a more user-centered approach, and should put a stronger emphasis on involving users' views and feedback. Studies have shown that developers consult other developers [7, 23] or with the Chief Privacy Offices (CPOs) [7, 8] when designing for privacy. However, keeping design in narrow "professional" circles is highly problematic. As danah boyd argues, it is crucial to understand the social and cultural factors involved in the context of the way systems are used: "technologists assume the most optimal solution is the best one, but this tends to ignore a whole bunch of social rituals that have value." [10].

Leaning only on internal testing before launching a new system or feature can end up in systems that mismatch users' privacy expectations. This is particularly important as end-users' privacy expectations are not only about the way their data is handled between them and the system (an aspect known as institutional privacy [51, 52]), but rather, expectations also relate to social privacy: how systems allow managing relationships between end-users, and the complexity that sharing and hiding information plays in these relationships [35, 51, 52]. Legal frameworks hardly address social privacy, as long as users have agreed to the terms of service [9]. However, consent does not necessarily mean that users' expectations are met, as can be evident in previous privacy designs that included consent but surprised users [20, 50].

To effectively receive feedback from end-users about their perceived privacy of the system, there is a need to reliably measure their observations. Many works have suggested methods and scales to measure people's privacy attitudes and concerns [15, 25, 28, 42, 43, 56, 60, 68]. Some of these studies have focused on systems' privacy evaluation. For example, Suh et al. have created a scale that measures users' burden in computing systems, which includes a specific construct to evaluate the system's privacy [60]. However, these studies have mainly dealt with institutional privacy, rather than social privacy [29, 30]. Our study extends this strand of research by working towards a tool that is built to measure how users perceive a particular design. Currently, there is no generic scale that can point to a feature that is considered as alarming and inappropriate by the end-users in any given system design.

In this study we attempted to develop and validate the reliability of a novel privacy scale that adds a social aspect which highlights the information flow between people using the system. We used the scale to explore whether the usage of A/B testing, also known as a controlled experiment, is applicable for privacy evaluation purposes.

We conducted a study with two major stages: 1) seeking to develop users' perceived privacy scale, and 2) comparing privacy designs by using the scale. We began with the scale development, in which we recruited 459 participants via Amazon Mechanical Turk (AMT). To validate the scale we used several methods including principal component analysis (PCA), exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). At the second stage of the study, we used the scale to compare two different privacy designs of given background scenarios, borrowing the controlled experiment methodology. We recruited 858 participants via AMT and found significant differences between the designs in three out of the five background scenarios. The study results show that a controlled experiment can be extended to privacy evaluation, mostly for social privacy. In the same manner they show that our scale is partially sensitive enough to differentiate between the two design variants, according to the differences in social and institutional information management and controls, as well as the overall risk users feel involved in using the system.

2. Background

2.1. Privacy by Design

Privacy, as defined by the sociologist Alan Westin, is "the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others." [63]. In contemporary information systems, the principle of control manifests itself in several ways. Specifically in the context of online social networks' (OSNs), several studies have found that there are two ways in which users think about privacy and their control over their data: 'institutional privacy', which reflects the data relationship between users and the system, and 'social privacy', which reflects the data relationship between users mediated by the system [35, 51, 52]. Raynes-Goldie has found that OSNs users are more concerned about controlling access to their information by other people, rather than about how companies are using the data [52].

Privacy-by-Design is an approach that advocates mitigating privacy threats from the very beginning of the system development, rather than by adding privacy-enhancing technologies after the fact [13, 38]. Recently, PbD is becoming a central tool in regulatory frameworks, endorsed by the U.S. FTC and the new European GDPR. The latter requires data controllers to implement "data protection by design and by default." [21].

Several studies have shown that when designing for privacy, developers mostly focus on security and protection against external entities [6, 23]. In the cases of encountering privacy issues, developers often see these as someone else's responsibility [23] or seek advice from other developers or legal and managerial entities inside their organizations [7]. Wong and Mulligan call to "bring design to the PbD table," enriching PbD practices and research with design as a conception of a process, rather than as a mere tool for implementing objectives [65]. We add to their call and focus on the users' aspect. We argue that PbD processes consistently neglect the end-user's perspective that should be considered during the development process along with efforts of compliance. Only a handful of PbD white-papers have recommended involving users or receiving feedback from them (the suggestion to use focus groups by the UK Information Commissioner's Office [27] is an exceptional example). Involving users is never a mandatory requirement in formal PbD processes, and there are no proven methods to carry out user feedback in scale.

2.2. Controlled Experiments in User Experience

Controlled experiment for evaluating user experience, popularly known as A/B testing, is a methodology that is used to have a better understanding of the advantages and disadvantages of different designs. In a controlled experiment, users are randomly exposed to one variant (for example, two different shapes of a button), in a persistent context, in which the difference between the variants is minimized. Referring to the previous example, the variants will differ only or mostly in the tested button [32]. Controlled experiments are essential tools for web-facing companies, using such experiments to gain valuable customers feedback in a short time. The experiments purposes are varying, including mostly monetization, and testing usability improvements [31]. In our study, we used a controlled experiment to compare privacy designs alternatives. We want to explore the applicability of using this methodology for privacy evaluation purposes, also noting that there are different privacy aspects, as social and institutional, which might behave differently.

2.3. Measuring Privacy Attitudes

Several scales that consider different aspects of privacy were developed over the years. However, these scales were mostly developed to measure individuals' privacy attitudes, as their personally privacy concerns, rather than to evaluate systems. One of the most used scales was developed by Malhotra et al., who developed the Internet Users' Information Privacy Concerns (IUIPC) scale [42]. Dinev et al. took a different perspective in which they referred to privacy as a state within a certain context. Therefore, rather than

measuring privacy concerns, they measured perceived (state of) privacy [15].

Previously mentioned scales measured users' privacy approaches, while not measuring privacy perceptions in the context of a specific system. Other studies developed scales that included constructs that measured system's privacy [28–30, 43, 60, 68]. However, these studies have focused on institutional privacy aspects, lacking the perspective of social privacy. Considering the substantial participation of users in social media, Page et al. called privacy researchers to refer to social privacy concerns [49]. Several studies in the networked privacy domain developed privacy scales that were specific to privacy in the context of social networks. Stutzman developed an instrument to elicit the users' attitudes about the access to personal information by different people with different relationships [59]. Young and Quan-Haase used their privacy protection strategies instrument [69] to find that Facebook users' have developed privacy protection strategies and that they are mostly used to protect users against social privacy threats and not against institutional privacy [69]. Considerable research was dedicated to understanding privacy in social networks [36, 44, 61]. However, as social privacy is a concern in every collaborative system, there is a need to understand user expectations regarding the way a given system allows users to manage information sharing and privacy between users.

2.4. Research Objectives

Taking the approach of Suh et al.s, in creating a scale for measuring user burden in systems [60], we aim to fill a gap in measuring the users' perceived privacy of a tested system. Moreover, it is unclear how various aspects of the system's privacy, including social and institutional privacies, affect users' perceived attitudes towards the system.

In this study, we aim to understand whether it is appropriate to use a controlled experiment, user-centered design methodology, to evaluate privacy design.

The first step was to define a reliable measurement, with which we can quantify the system's privacy, as it perceived by the users. Our literature review had brought us to investigate two distinct privacy aspects:

H1. Users' perceived privacy of a given system consists of two distinct aspects: social privacy and institutional privacy.

Once we have a reliable measure, we can explore whether controlled experiment methodology is applicable to compare privacy designs:

RQ.1. Can controlled experiment methodology differentiate between privacy designs?

RQ.2. Does the controlled experiment methodology applicability depend on the tested privacy aspect (social or institutional)?

3. Initial Scale Design

The goal of our scale is to measure end-users' perceived privacy of a specific information system, as we named it: Users' Perceived Systems' Privacy (UPSP) scale. We strongly based our scale on previous studies that created privacy scales. Some of the studies presented general privacy scales [4, 15, 19, 25, 42, 56] and others were specified to privacy in the context of OSNs [36, 44, 59, 61, 69]. Based on the literature review we identified a gap of a missing scale to measure perceived privacy from a social aspect, and that is aimed to evaluate an information system. Therefore, we attempted to create a scale that covers simultaneously both institutional-related aspect, which refers to privacy aspects between the user and the system, and social-related aspect, which refers to privacy aspects between the user and other people.

At the first stage of the scale development, we created a list of questions to represent institutional-related aspect. We chose several constructs that appeared on the previous general (i.e., not OSNs specified) privacy questionnaires and made adaptations when required, to represent questions about users' perceived privacy of the system. The chosen constructs were: perceived information control, confidentiality, importance of information transparency, secondary usage, data deletion, perceived privacy risk, and information sensitivity.

At the second stage, we developed new social-related questions based on the previously mentioned constructs and based on OSNs' specified privacy questionnaires. The social-related questions included two additional constructs, according to the original study upon which the questions are based on protection strategies [69] and identity sharing [59]. See Appendix A for the final questionnaire questions and their original constructs. Finally, our preliminary questionnaire included a set of 47 questions. Twenty-seven questions were institutional-related, and 20 questions were social-related.

3.1. Experimental Design and Recruitment

To evaluate our scale, we recruited participants via Amazon Mechanical Turk (AMT). Redmiles et al. found that MTurk responses regarding security and privacy aspects can be generalized to a broader population [53]. Our scale was aimed to assess users' perception of an information system, similar to Suh et al. [60]. To ensure generalization, we tested our scale while referring to several systems, but each participant was exposed to one system only. The systems we chose were Facebook, YouTube, and WhatsApp. Two of the systems have a prominent social aspect, which may raise

privacy concerns (Facebook, WhatsApp), and a third system has a smaller social aspect (YouTube), to cover varying systems.

We screened the participants in several ways. They were required to be 18 years of age or older, and to use the systems frequently (approximately at least once a week). From AMT perspective, the participants were based in the U.S., had an approval rate of 95% or greater, and had at least 100 HITs approved. As we intended to do exploratory factor analysis, we recruited 300 participants. Bryant and Yarnold suggested a minimum ratio of 1:5 of participants per items to conduct EFA [11]. Our questionnaire included 47 items. Thus we assumed we would have at least the desired minimum if recruiting 300 participants. The participants were randomly assigned to one of the systems only. The questions were presented as statements, and the participants were asked about the extent to which they agree with each statement. We used a seven-point Likert scale, where 1 represented low agreement and 7 represented high agreement. The two sub-scales, institutional-related and social-related, were randomly ordered, and the questions within the subscales were randomly ordered as well. We gave participants an "I do not know" option so that we could determine which questions were problematic. The entire study, including all three experiments, was authorized by the institutional ethics review board (IRB) and occurred between May 2018 and February 2019.

Qualified participants followed a link that randomly assigned each participant to one of the three links to the questionnaire, each referring to one of the systems (Facebook, YouTube, WhatsApp). Following previously developed privacy and usability scales [15, 60], we did not include reversely coded statements. In these scales reversely coded questions are rare due to the added complexity they add to the scale. The questionnaire was built using the Qualtrics commercial web survey service. The participants completed an IRB-approved consent form on participation limitations. The mean completion duration was approximately 6.5 minutes, and we paid $0.4 per assignment completion.

Similar to the methods used by Egelman and Peer [17], we took two steps to mitigate social desirability bias on participants' responses, in which some participants may answer questions according to what they believe to be viewed as favorably by others [14]. First, we did not mention "privacy" during recruitment to minimize selection bias. Second, we asked all participants to complete the 10-item Strahan-Gerbasi version of the Marlowe-Crowne Social Desirability Scale [57], which we then correlated with participants' responses to our survey questions. We checked for the existence of straight-lining behavior, in which a participant answers the same answer for all the questions, generally considered to point at superficial thinking [70]. Lastly, following Goodman et al.'s [48] study on AMT, we phrased screening questions to identify participants who would not follow the survey's instructions. If participants failed to answer both questions incorrectly, we excluded their records. After filtering out participants who completed the screening task incorrectly (n = 59) and checking for straight lining behavior (n = 0), we were left with 241 valid responses. See Appendix C for the screening task questions. See Appendix B for the participants' age and gender distribution. The group size of each system was: Facebook: 67, WhatsApp: 78, YouTube: 96.

3.2. Results

We performed Pearson correlations between the Strahan-Gerbasi social desirability scale and each question. Except for one question, the observed Pearson's r values corresponded to less than 5% common variance. The remaining question corresponded to less than 10% common variance, the cutoff of which one relationship represents practical importance. Therefore, we chose to treat all the questions as lacking social desirability bias. The result suggests that participants answered truthfully and consistently.

We proceeded to perform Exploratory Factor Analysis (EFA) using Promax rotation to determine which questions should remain in the final questionnaire [58]. We performed four EFAs: one analysis included all the systems and three others for each system separately. We used a loading value of 0.5 as a cut off to include the item within the questionnaire, similar to Egelman and Peer [17]. The number of factors we extracted for each EFA was based on Principal Component Analysis (PCA), using a parallel analysis [2]. For all-systems and Facebook EFAs we extracted four factors, for WhatsApp and YouTube we extracted three and two factors, respectively. First, we removed questions that were below the cutoff value in the EFA that referred to all the systems (7 items). Next, we removed the questions that were below the cutoff in the EFA of each specific system (12 items). We looked for questions that will fit as much as possible to varying types of systems. Therefore, if a question was not good enough for a certain type of system, but was with an appropriate loading value in the other systems, we chose to eliminate it.

Lastly, we re-run the EFA with the remaining 28 questions using the responses of all the systems, ensuring that all the items' loadings are above the cutoff. At this point, we extracted three factors according to the parallel analysis and this analysis resulted also in a 28 items questionnaire. We also checked that none of the final questions was extremely problematic regarding the number of participants choosing "I do not know." Among all the questions (47 items) the highest rate of the N/As responses was 9.5%, and the mean rate was 3.5%. Among the final set of questions, the highest

rate was 6.2% and the mean rate was 3.6% Therefore, we kept all 28 questions.

Finally, our analyses yielded three factors, which we named as institutional, social and risk, partially confirming our hypothesis. Our results showed that users' perceived privacy consist of three distinct aspects, and not only of institutional and social aspects. The questions of the institutional factor are taken from our initial institutional-related section. Respectively, the questions of the social factor are taken from our initial social-related section. However, the questions of the risk factor are mixed of the two original sections, and they are all referring to risk or information sensitivity.

4. Finalizing the Scale

We recruited an additional cohort of participants so that we could perform a Confirmatory Factor Analysis (CFA) [58] on the reduced questionnaire. The participants had answered 33 items questionnaire, based on EFA using Varimax rotation. Further Promax rotation eventually reduced the number to of questions to 27.

4.1. Method and Demographics

We recruited 300 new participants to respond to the set of the chosen questions. Since at this stage we aimed to have a final scale, we preferred to have more participants per items. Therefore we recruited 300 participants, despite the reduction of the total items number. Following our preliminary results, we removed the Strahan-Gerbasi scale. We kept our screening questions to allow us the removal of suspicious careless responses. We removed the option to answer "I do not know." The course of the experiment was similar to the former experiment. The mean completion duration was approximately 4.23 minutes, and we paid $0.4 per assignment completion. After filtering out participants who completed the screening task incorrectly (n = 82) and checking for straight lining behavior (n = 4), we were left with 214 valid responses. See Appendix B for the participants' age and gender distribution. The group size of each system was: Facebook: 89, WhatsApp: 52, YouTube: 73.

4.2. Results

In the following section, we describe several heuristics aimed to explore our scale validity [58]. First analyses are aimed to ensure constructs validity, using PCA, EFA, CFA, as well as convergent and discriminant validity. Next, we performed a reliability analysis, using Cronbach's alpha analysis. The constructs and reliability analyses were conducted based on all the responses (n = 214). Lastly, we analyzed the scale sensitivity, in which we compared the three systems. The sensitivity analysis resulted in changing some of the questions wordings, as we describe in the coming paragraphs.

5.2.1 Constructs Validity. First, we performed a PCA using parallel analysis to extract the number of factors [2]. The scree plot pointed to three factors, as expected. Next, we performed an EFA using Promax rotation and considered an item to be loaded on a factor if its loading exceeded 0.5. The factors and the questions within them were the same as in the preliminary scale. Therefore, all the 27 questions remain within the final scale. The three factors that we extracted predicted 56.1% of the variance. The themes of the factors remained the same: institutional, risk and social. Each of these factors accounted for 25.7%, 16%, and 14.4% of the variance, respectively.

Next, to validate our EFA results, we performed a CFA and examined the model's goodness-of-fit. Multiple popular metrics showed that our data supported the model. Our relative chi-square statistic, $\chi 2/df$, was 2.0. There is no consensus regarding an acceptable cutoff for the ratio, and recommendations range from 5.0 to 2.0. Therefore, our result is acceptable [26]. Our analysis yielded Root Mean Square Error of Approximation (RMSEA) of 0.068 and a Standardized Root Mean Square Residual (SRMR) of 0.065, which is following the recommended maximum cutoff point of 0.08 for both measures [26]; Finally, our Comparative Fit Index (CFI) was 0.92 and Tucker-Lewis Index (TLI) was 0.91, which are above the recommended cutoff of 0.90 [45].

Finally, we conducted convergent and discriminant validity tests. Convergent validity ensures sufficient inter-correlation between each of the construct's variables, while discriminant validity ensures that the constructs are distinct [58]. We found that the average variance extracted (AVE) of each factor is above the acceptable cut-off of 0.5, pointing to convergent validity [24]. As per discriminant validity, we found that the square root of the AVE of each construct was greater than the correlations between the construct and the other constructs in the model [24]. The results are summarized in Appendix E.

4.2.2 Reliability Analysis. We examined the scale reliability using Cronbach's alpha. The computed Cronbach's alpha for the full scale was 0.95. Next, all of subscales had excellent internal consistency as well (> 0.9) [22]: institutional: α = 0.95, social: α = 0.9, and risk: α = 0.9. Thus, we concluded that our full scale and the sub-scales each had high reliability.

4.2.3 Scale Sensitivity Analysis. Lastly, we compared the systems using the new scale, to have a preliminary notion whether the scale is sensitive enough to detect differences in perceived privacy between systems, similar to the approach taken by Suh et al. [60]. First, we averaged each scale per participant, so each participant now had three scores (institutional, social, risk). Next, we performed Analysis of Variance (ANOVA) per each sub-scale, in which we tested whether there is a significant difference between the sys-

tems. We performed a Tukey post-hoc analysis to find between which systems the difference in the mean score of the scale was significant. The results are summarized in table 1. We can see that for both scales, institutional and risk, there were significant differences between some of the systems. We were surprised by the results, since we would aspect to see a difference in the social aspect primarily, at least between Facebook and YouTube or WhatsApp and YouTube since we chose the systems based on their social aspect.

The ANOVA and the Tukey analyses results brought us to reconsider the statements wordings. The social statements were completely developed and phrased by us, while we considered the previous literature in mind. The results had brought us to notice that we did not include the specified name of the relevant system almost in all social questions, unlike in the other sub-scales questions, which we only slightly modified previously developed questions. Therefore, we added the specified name of the system for those statements as well. To conclude, we see that the survey was sensitive to an extent at this point, detecting some differences between different systems, before finalizing the social statements wordings. See Appendix A for the final suggested scale.

Table 1. Comparing the systems (Facebook, WhatsApp, and YouTube), exploring in which subscales there are significant differences in the mean score.

	ANOVA	Systems compared	Adj. *p*-value (Tukey)
Institu.	**$F(2,211) = 5.09, p < 0.01$**	**WA-FB**	**0.007**
		YT-FB	0.09
		YT-WA	0.52
Social	$F(2,211) = 1.69, p = 0.19$	WA-FB	0.67
		YT-FB	0.492
		YT-WA	0.168
Risk	**$F(2,211) = 8.63, p < 0.01$**	**WA-FB**	**0.001**
		YT-FB	**0.002**
		YT-WA	0.885

5. Controlled Experiment and Using the Scale

In the previous sections, we described the development and the steps we took to ensure the internal validation of our scale. In this section, we describe how we used the scale to answer our research questions referring to the applicability of controlled experiment to privacy purposes evaluation, and in which circumstance it can be applied. Unlike as with the previous sections, in which we compared between real systems, and therefore were unable to control for different variables related to privacy design, in this experiment we created scenarios and controlled the desired variables.

5.1. Method

To answer our research questions, we designed a between-subject user study, using an online experiment that included a scenario presentation followed by the UPSP scale. We created five scenarios, and per each scenario we created two cases, differing in their privacy design: privacy intrusive design versus privacy respective design. Altogether, we had five background scenarios and ten cases. We recruited 1,026 participants, and they were randomly assigned to one of the ten scenario-case combinations only. We used G*power to estimate the required sample size for T-test analyses and found that the required sample size is 88 participants per group (effect size $d = 0.5$, $\alpha = 0.05$, $1-\beta = 0.95$) [18]. Therefore, we recruited 100 participants per case, and also run several pilots to make sure that the experiments work well, eventually recruited 1,026 participants. Screening parameters for recruiting participants were similar to previous experiments (Sections 3 and 4), except they were not required to be Facebook, WhatsApp or YouTube frequent users. In this experiment, we changed the screening task by shortening the paragraph the participants were required to read (Appendix D).

The background scenarios were developed based on similar principles of previously real privacy case studies that had occurred. For example, one of the scenarios was similar to WhatsApp status update and referred to privacy concerns that were raised as a result of launching the feature [3, 62, 64]. As for the visualization perspective, we designed the general scenarios and the cases based on Ayalon and Toch study [5]. They found that when presenting the privacy characteristics of a system, there is a need to show the human aspect of the problem, rather than presenting it only as a matter of data flow. Qualified participants were first presented with a general explanation, in which the participants were informed that they are about to read a description of a future app and that they are asked to imagine themselves as users in the specific scenario. Next, the participants were presented with the case details, which was consisted of four information sections: 1) *App Presentation* – the app's name followed by a very short description. If required, additional information about the app was provided; 2) *App demonstration* - screenshot, one or more, demonstrating some of the app's interfaces; 3) *Feature presentation* (optional) – in case of a feature within an app, specific information about the feature was provided; 4) *Case description* - description of the specific case and a relevant screenshot, one or more. Lastly, the participants were presented with the

UPSP scale questions. The statements were presented as three sub-scales: institutional, social, and risk. The subscales were ordered accordingly, and the statements within each subscale were randomly ordered.

As we were interested in testing the different privacy aspects of information systems, three background scenarios had a prominent social aspect, and two background scenarios were focused on the institutional aspect. The applications' names that were used as the background scenarios were invented, but we have based the applications' functionalities on existing applications. The three social applications and features used were: 1) iFindRest, which helps the users with finding restaurants based on their location and reserving a table; 2) Message4All app, Tale feature. The app is a messenger app, and the feature enables the users to show content to all the apps' users who have the user's phone number, for a limited time; 3) Message4All app, focusing on groups' details disclosure. Users can view their contacts' shared and non-shared groups. The remaining two institutional applications used were: 4) iFit, a fitness app which helps the users with doing exercises; 5) Message4All app, ads publications, in which ad appears in the chat interface. See Appendix F to view the different scenarios and the two cases per each scenario.

Taking all the participants' responses across the scenarios, the mean completion duration was approximately 6.7 minutes, and we paid an average of $0.63 per assignment completion. After filtering out participants who completed the screening task incorrectly (n = 160) and checking for straight lining behavior (n = 8), we were left with 858 valid responses. See Appendix B for the participants' age and gender distribution. The group size of each scenario-case combination was: iFindRest: intrusive 96, respective 76; Message4All - Tale: intrusive 80, respective 76; Message4All - Groups: intrusive 99, respective 77; iFit: intrusive 86, respective 79; Message4All - Ad: intrusive 87, respective 102.

5.2. Results

We began with re-validating our scale using CFA. Based on the entire sample (n = 858), we examined the model's goodness-of-fit using the same fit statistics as previously and found that our data supported the model: $\chi 2/df = 5.29$, RMSEA = 0.071, SRMR = 0.049, CFI = 0.94, TLI = 0.93. Next, as we assured we could use the scale, we turned to compare between the two cases (intrusive vs. respective) per each scenario. First, we averaged each scale per participant to create three distinct scores (institutional, social, risk). We wanted to compare the two cases per each sub-scale. Therefore, we performed T-tests and used Bonferroni correction for multiple comparisons, in which the *p*-values were multiplied by the number of comparisons.

Our results showed that within different scenarios the differences between the cases were significant and are summarized in Table 2. For all the social background scenarios, we found a significant difference between the cases for at least one of the sub-scales (institutional, social, risk). In the scenario that referred to iFindrest we found that the intrusive design was perceived as riskier compared to the respective design (p = 0.045), and we did not find significant differences in the other categories. In the Message4All app that referred to the Tale feature, we found significant differences between the cases for two of the subscales ($p < 0.05$). The privacy respective design was perceived as respective from the institutional and social aspects. Surprisingly, in the Message4All app that referred to groups information disclosure we found that the respective design was considered as riskier compared to the intrusive design ($p = 0.023$). For the other categories, the difference was insignificant. However, in the institutional background scenarios, iFit and Message4All with the ad presentation, we did not find significant differences between the cases for any of the sub-scales. Figure 1 summarizes the mean sub-scales scores of each scenario, comparing the two cases.

Table 2. Comparing the cases per each scenario, exploring in which subscales there are significant differences in the mean score.

Scenario	Sub-scale	Res.	Int.	Adj. *p* value	Cohen's d
Social background scenarios					
iFindRest	instit.	3.88	3.73	1	0.10
	social	4.11	4.33	0.301	0.26
	risk	**4.18**	**4.49**	**0.045**	**0.38**
Message4-All, Tale	**instit.**	**4.45**	**3.73**	**0.003**	**0.54**
	social	**4.72**	**4.40**	**0.018**	**0.45**
	risk	4.77	4.6	0.32	0.26
Message4-All, Group	instit.	4.04	3.76	0.689	0.18
	social	4.81	4.46	0.051	0.37
	risk	**4.97**	**4.64**	**0.023**	**0.41**
Institutional background scenarios					
iFit	instit.	3.34	3.07	0.756	0.18
	social	3.84	3.84	1	0.0
	risk	4.05	4.01	1	0.06
Message4-All, Ad	instit.	3.67	3.33	0.523	0.20
	social	4.43	4.44	1	0.01
	risk	4.68	4.77	1	0.10

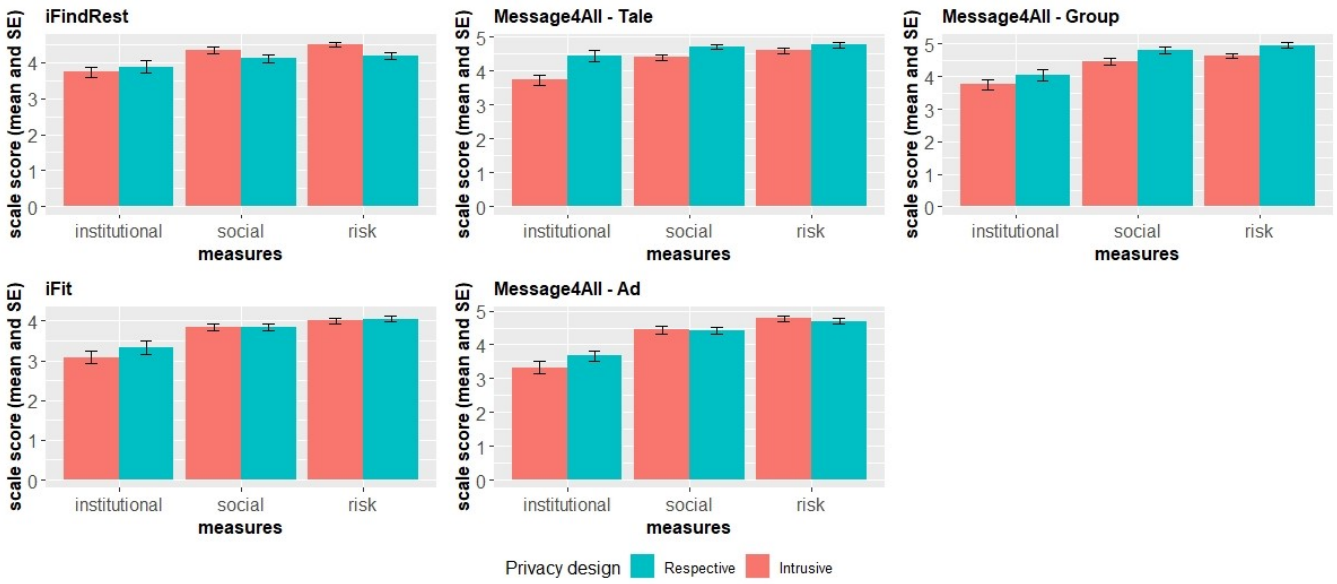


Figure 1. Per each scenario, we compared the two privacy designs, respective vs. intrusive, per each subscale. For example, in the Message4All - Tale scenario, there were significant differences between the designs for two constructs: institutional and social.

6. Discussion

This study aimed to explore the applicability of controlled experiment methodology to evaluate privacy designs. Towards our exploration, we first attempt to developed a measure to quantify users' privacy perception of a given information system. The scale, in its current form, shows that users perceive information system's privacy via three distinct aspects: institutional, social and risk. This result partially confirms our hypothesis, which referred to institutional and social aspects only. Using our scale, we compared designs which differed in the extent to which they were privacy intrusively designed. Our findings point to a limited ability of controlled experiment methodology to serve as a sensitive way to evaluate privacy design. We saw that the differences between the designs received greater attention when the demonstrated privacy issue had a prominent social aspect, and not, for example, an institutional aspect.

6.1. Theoretical Implications

We were motivated by the Privacy by Design (PbD) approach and encouraged by the inclusion of PbD in the European GDPR in 2018. However, PbD can be criticized in a similar way that mainstream system design was criticized by the User Centered Design approach [39]. We argue that ignoring the users and focusing on compliance to regulation will result in systems that are legal but would still make users uncomfortable and go against social norms in particular contexts [46]. Our results point to particular contexts in which system design can be considered as inappropriate. Specifically, our findings suggest that privacy issues with a salient social aspect were highly prone to criticism and observation by the users, compared to the institutional aspect.

Involving the users in the design process may lead to surprising, sometimes even paradoxical, results. In the Message4All scenario, the design that included a message that reminded users that they can disable the disclosure of sensitive information was considered riskier than the alternative design that did not included a message (but in which the sensitive information was collected). Knijnenburg and Kobsa reported on similar results in which messages that were aimed to justify information disclosure decreased the users' trust and satisfaction of the tested system [29]. This result highlights the need to involve the users, showing that the designers, in this case the papers' authors, cannot fully estimate users' perceptions and understandings without asking them directly.

Our findings highlight the promises, and limitations, of our methodology. Controlled experiment methodology is widely used today to provide a fast and affordable evaluation of computing systems. The widespread deployment of this methodology demonstrates that some aspects of user-centered design (UCD) approach are becoming well accepted by today's computing systems' developers, designers and anyone who is part of the decision-making process.

Investigating the applicability of the scale to privacy design evaluation revealed a more complicated picture, in which we saw a significant difference between the cases only in some of the background scenarios. There are several possible ex-

planations for the different results between institutional and social scenarios. One of the explanations can be the difference between the systems' *privacy affordances* [37, 40, 55]. General perceived affordances refer to both the perceived and actual properties of a certain "thing" that define how it can be used [47]. Referring to privacy, previous studies referred to privacy affordances in several contexts. For example, Kou et al. found that Facebook's features as chatrooms and posts' privacy settings affect the users' self-presentation behavior [37]. Liebling and Preibusch suggested to improve gaze tracker by adding privacy affordances to increase the users' privacy [40]. In the current paper we refer to privacy affordances as the ease of the users' ability to understand or foresee the possible consequence of a given privacy issue.

Privacy affordances, as raised in our results, can add another perspective to the privacy paradox debate. The Privacy Paradox is a term usually referring to the gap between people's stated privacy concerns (high) and their actual behavior (disclosing a large amount of information, for example) [33]. Many studies are exploring the paradox, some suggesting possible explanations. One type of explanations refers to the users' constraints of bounded rationality and incomplete information [1], and information asymmetries [12]. These explanations are referring to the users' limited knowledge of the possible consequences of their privacy-related behavior. Our results support these explanations, pointing to different privacy affordances in different types of privacy aspects. For social aspects, privacy affordances are straightforward allowing users to easily imagine possible consequences. As users are actively engaging with social applications, serving as both publishers and audience, users understand what could be the results of posting information to their entire contact list. On the other hand, as with institutional aspects, privacy affordances are much weaker. It is harder to understand the complicated information flows that are behind the way contemporary platforms collect and process their personal information, and which other unknown institutions might access their information and use it as well.

Methodological explanations to the sensitivity of the scale are possible as well. First, the experiment consisted of no more than five scenarios. Possibly, the designs of the institutional scenarios (Message4All – Ad, and iFit) were not sufficiently different surface the problematic privacy issues they ought to represent. Perhaps, if we had used other institutional scenarios we would have received different results. Second, although the scale was validated for its reliability using several acceptable methods, further exploration and improvement is required. Performing EFA had brought us to conclude that there are three distinct constructs. However, it is possible that the difference between the construct "risk" and the two other constructs is not big enough, thus influencing on the ability to differentiate between the privacy designs.

6.2. Using the Scale and Design Implications

In this study, we have attempted to develop a scale to measure systems' privacy. Although the scale was designed to capture the users' perceived privacy of a specific system, without limiting the type or the context of the system, the results point to the scale's partial success in fulfilling its intended role. We suggest possible implementations of the scale, however, not without mentioning its limitations to differentiate between all privacy designs. Future implementation of the scale should consider its possible inability to differentiate between privacy designs that are lacking of social aspect.

Following our first suggested explanation, privacy affordances, beyond its theoretical contribution, it also has practical implications. The UPSP scale aims to provide knowledge about the users' perceptions of a system's privacy. Finding significant differences between the designs can point to a good usage of privacy affordances while lacking differences can highlight that the users might not fully understand the possible privacy consequences. Systems' developers should not necessarily give themselves a pat on the back when they do not find a significant difference between the system's privacy designs. They should first look at the score, whether pointing to a high sense of perceived risk, for example. In addition, if in both cases perceived risk is high, but they do not significantly differ, the developers should consider the option the users simply cannot imagine what might be the results of their privacy behavior.

Controlled experiments provide practitioners with new knowledge, for example, which design resulted in a higher conversion rate [32]. Using the UPSP scale provides new knowledge as well. The novelty of our scale is its multi-facets, covering distinct privacy aspects (social, institutional and risk), and its approach, aimed to evaluate systems, rather than individuals' attitudes, as their general privacy concerns. While considering the scale's current uncertain ability to differentiate between privacy designs with a prominent institutional aspect, information system's developers can benefit from using our scale in several ways. First, the scale itself, resulting in three distinct scores per each tested design. A controlled experiment on its own will not provide the required understanding. For example, if the conversion rate was used as a measure, the developers would still lack the knowledge of what was wrong, or right, as perceived by the users. Second, the scale brings the users' perceptions, which might differ from the developers' perception and even from privacy experts. This is similar to other fields as user experience, user interface, usability, and others. Experts are required to set the hypotheses, but the users will eventually determine whether to confirm or reject them. Third, in their study Spiekermann and Cranor suggested guidelines for building privacy-friendly systems, distinguishing between

"privacy-by-policy" versus "privacy-by-architecture." [54] Our study results suggest adding more spheres that should be considered, especially with the rise of social privacy aspect since their study was conducted.

The last implication for design is our structured suggested framework for evaluating users' perceived privacy, as was described in section 5.1. The framework is necessary to demonstrate privacy issues simply and concisely, and yet, understandable by the general population. The process includes five steps: general scenario level: 1) App Presentation; 2) App demonstration; 3) Feature presentation (optional); different versions level: 4) Case description. 5) Lastly, answering the UPSP scale.

6.3. Limitations and Future Work

Our study is subject to several limitations that impact its applicability for design and research. First, to evaluate our scale we used five background scenarios. While we have strived to base the scenarios on typical privacy designs, further studies and practical experience are necessary to evaluate it the real world. Second, the participants reflected their opinion about the presented scenario. Their actual behavior in the context of a similar incident might differ, possibly reflecting a weaker difference between the cases. Third, as norms around privacy evolve these days quickly, the scale should be continuously evaluated to see that it reflect contemporary notions. Lastly, as we have suggested a method to evaluate privacy designs, the study population should be sampled and adjusted to particular systems and scenarios. As with many privacy studies, the use of Mechanical Turk as the study's population may not reflect the actual demographics of the intended system.

Based on the study results we are developing *A/P(rivacy) Testing*[1], a platform that will enable other researchers and developers to use our scale and to compare privacy designs easily. Future studies can explore real systems or focus on specific challenges, for example, exploring different ways to visualize consent form and the visualization's effect on users' perceived privacy.

7. Acknowledgment

This work was supported by the Shulamit Aloni Scholarship by the Israeli Ministry of Science and Technology, grant number 314575, and by the ICRC – Blavatnik Interdisciplinary Cyber Research Center, grant number 590713. We would also like to thank Luiza Jarovsky for helping us with finalizing the scale and Shany Peter for developing the A/P Testing tool.

References

[1] Acquisti, A. and Grossklags, J. 2005. Privacy and rationality in individual decision making. *IEEE Security and Privacy*. 3, 1 (2005), 26–33.

[2] Ahmad, sarah sabir 1999. Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*. 4, 3 (1999), 272.

[3] As Facebook Raised a Privacy Wall, It Carved an Opening for Tech Giants: 2018. *https://www.nytimes.com/2018/12/18/technology/facebook-privacy.html?module=inline*.

[4] Awad, N. and Krishnan, M. 2006. The personalization privacy paradox: an empirical evaluation of information transparency and the willingness to be profiled online for personalization. *Mis Quarterly*. 30, 1 (2006), 13–28.

[5] Ayalon, O. and Toch, E. 2018. Crowdsourcing Privacy Design Critique : An Empirical Evaluation of Framing Effects. *Submitted*. (2018).

[6] Ayalon, O., Toch, E., Hadar, I. and Birnhack, M. 2017. How Developers Make Design Decisions about Users' Privacy: The Place of Professional Communities and Organizational Climate. *Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing - CSCW '17 Companion*. (2017), 135–138.

[7] Balebako, R., Marsh, A., Lin, J., Hong, J.I., Cranor, L.F. and Faith Cranor, L. 2014. The Privacy and Security Behaviors of Smartphone App Developers. *Workshop on Usable Security (USEC)*. (2014).

[8] Bamberger, K.A. and Mulligan, D.K. 2011. *Privacy on the Books and on the Ground.* MIT Press.

[9] Bechmann, A. 2014. Non-Informed Consent Cultures: Privacy Policies and App Contracts on Facebook. *Journal of Media Business Studies*. 11, 1 (2014), 21–38.

[10] Boyd, D. 2010. Making Sense of Privacy and Publicity. *South by Southwest (SXSW 2010)– transcription of the talk*.

[11] Bryant, F.. B. and Yarnold, P.. R. 1995. Principal-components analysis and exploratory and confirmatory factor analysis. *Reading and understanding multivariate statistics*. L.G. Grimm and P.. R. Yarnold, eds. American Psychological Association. 99–136.

[12] Buck, C., Horbel, C., Germelmann, C.C. and Eymann, T. 2014. The unconscious app consumer: Discovering and comparing the information-seeking patterns among mobile application consumers. *European Conference on Information Systems (ECIS)* (2014).

[13] Cavoukian, A. 2009. Privacy by design: The 7

[1] www.aprivacytesting.com

foundational principles. Information and Privacy Commissioner of Ontario, Canada.

[14] Crowne, D.P. and Marlowe, D. 1960. A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*. 24, 4 (1960), 349–354.

[15] Dinev, T., Xu, H., Smith, J.H. and Hart, P. 2013. Information privacy and correlates: An empirical attempt to bridge and distinguish privacyrelated concepts. *European Journal of Information Systems*. 22, 3 (2013), 295–316.

[16] Egelman, S., Felt, A.P. and Wagner, D. 2013. Choice architecture and smartphone privacy: There's a price for that. *The Economics of Information Security and Privacy*. (2013), 211–236.

[17] Egelman, S. and Peer, E. 2015. Scaling the Security Wall: Developing a Security Behavior Intentions Scale (SeBIS). *Proceedings of the ACM CHI'15 Conference on Human Factors in Computing Systems*. 1, (2015), 2873–2882.

[18] Faul, F., Erdfelder, E., Lang, A.G. and Buchner, A. 2007. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*. 39, 2 (2007), 175–191.

[19] Featherman, M.S. and Pavlou, P.A. 2003. Predicting e-services adoption: A perceived risk facets perspective. *International Journal of Human Computer Studies*. 59, 4 (2003), 451–474.

[20] Felt, A.P., Egelman, S. and Wagner, D. 2012. I've got 99 problems, but vibration ain't one. *Proceedings of the second ACM workshop on Security and privacy in smartphones and mobile devices - SPSM '12*. (2012), 33.

[21] GDPR: *https://gdpr-info.eu/art-25-gdpr/*. Accessed: 2018-01-16.

[22] George, D. and Mallery, P. 1999. *SPSS for Windows Step by Step: A simple guide and reference*. Needham Heights, MA: Allyn & Bacon.

[23] Hadar, I., Hasson, T., Ayalon, O., Toch, E., Birnhack, M., Sherman, S. and Balissa, A. 2017. Privacy by designers: software developers' privacy mindset. *Empirical Software Engineering*. (2017), 1–31.

[24] Hair, J.F., Black, W.C., Babin, B.J. and Anderson, R.E. 2014. *Pearson New International Edition: Multivariat Data Analysis*.

[25] Hong, W. and Thong, J.Y.L. 2013. Internet privacy concerns: an integrated conceptualization and four empirical studies. *Mis Quarterly*. 37, 1 (2013), 1–3.

[26] Hooper, D., Coughlan, J., Mullen, M.R., Mullen, J., Hooper, D., Coughlan, J. and Mullen, M.R. 2008. " Structural Equation Modelling: Guidelines for Determining Model Fit Structural Equation Modelling: Guidelines for Determining Model Fit. *The Electronic Journal of Business Research Methods*. 6, 1 (2008), 53–60.

[27] ICO (Information Commissioner's Office) 2014. Conducting privacy impact assessments code of practice. *Ico.Org.Uk*. (2014), 1–55.

[28] Jarvenpaa, S.L., Tractinsky, N. and Saarinen, L. 1999. Consumer Trust in an Internet Store: a Cross-Cultural Validation. *Journal of Computer-Mediated Communication*. 5, 2 (1999).

[29] Knijnenburg, B.P.B. and Kobsa, A. 2013. Making decisions about privacy: Information disclosure in context-aware recommender systems. *ACM Transactions on Interactive Intelligent Systems*. 3, 3 (2013), 20:1–20:23.

[30] Kobsa, A., Hichang, C. and Knijnenburg, B.P. 2016. The Effect of Personalization Provider Characteristics on Privacy Attitudes and Behaviors: An Elaboration Likelihood Model Approach Alfred. *Journal of the Association for Information Science and Technology*. 67, 11 (2016), 2587–2606.

[31] Kohavi, R., Deng, A., Frasca, B., Walker, T., Xu, Y. and Pohlmann, N. 2013. Online controlled experiments at large scale. *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '13*. (2013), 1168.

[32] Kohavi, R., Longbotham, R., Sommerfield, D. and Henne, R.M. 2009. Controlled experiments on the web: Survey and practical guide. *Data Mining and Knowledge Discovery*. 18, 1 (2009), 140–181.

[33] Kokolakis, S. 2017. Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. *Computers and Security*. 64, (2017), 122–134.

[34] Koops, B.J. and Leenes, R. 2014. Privacy regulation cannot be hardcoded. A critical comment on the "privacy by design" provision in data-protection law. *International Review of Law, Computers & Technology*. 28, 2 (2014), 159–171.

[35] Krasnova, H., Günther, O., Spiekermann, S. and Koroleva, K. 2009. Privacy concerns and identity in online social networks. *Identity in the Information Society*. 2, 1 (2009), 39–63.

[36] Krasnova, H., Spiekermann, S., Koroleva, K. and Hildebrand, T. 2010. Online social networks: Why we disclose. *Journal of Information Technology*. 25, 2 (2010), 109–125.

[37] Kuo, F.-Y., Tseng, C.-Y., Tseng, F.-C. and Lin, C.S.

2013. A Study of Social Information Control Affordances and Gender Difference in Facebook Self-Presentation. *Cyberpsychology, Behavior, and Social Networking*. 16, 9 (2013), 635–644.

[38] Langheinrich, M. 2001. Privacy by Design - Principles of Privacy-Aware Ubiquitous Systems. *3rd international conference on Ubiquitous Computing*. (2001), 273–291.

[39] Law, E.L.-C., Roto, V., Hassenzahl, M., Vermeeren, A.P.O.S. and Kort, J. 2009. Understanding, scoping and defining user experience. *Proceedings of the 27th international conference on Human factors in computing systems - CHI 09*. June 2014 (2009), 719.

[40] Liebling, D.J. 2014. Privacy Considerations for a Pervasive Eye Tracking World. (2014), 1169–1177.

[41] Lin, J., Sadeh, N., Amini, S., Lindqvist, J., Hong, J.I. and Zhang, J. 2012. Expectation and purpose: understanding users' mental models of mobile app privacy through crowdsourcing. *Proceedings of the 2012 ACM Conference on Ubiquitous Computing - UbiComp '12*. (2012), 501.

[42] Malhotra, N.K., Kim, S.S. and Agarwal, J. 2004. Internet users' information privacy concerns (IUIPC): The construct, the scale, and a causal model. *Information Systems Research*. 15, 4 (2004), 336–355.

[43] Metzger, M.J. 2004. Privacy, Trust, and Disclosure: Exploring Barriers to Electronic Commerce. *Journal of Computer-Mediated Communication*. 9, 4 (2004).

[44] Mohamed, N. and Ahmad, I.H. 2012. Information privacy concerns, antecedents and privacy measure use in social networking sites: Evidence from Malaysia. *Computers in Human Behavior*. 28, 6 (2012), 2366–2375.

[45] Netemeyer, R.G., Bearden, W.O. and Subhash, S. 2003. *Scaling procedures: Issues and applications*. Sage Publications.

[46] Nissenbaum, H. 2010. *Privacy in context: Technology, policy, and the integrity of social life*. Stanford University Press.

[47] Norman, D. 2013. *The design of everyday things: Revised and expanded edition*. Basic books.

[48] Oppenheimer, D.M., Meyvis, T. and Davidenko, N. 2009. Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*. 45, 4 (2009), 867–872.

[49] Page, X., Tang, K., Stutzman, F. and Lampinen, A. 2013. Measuring networked social privacy. *Proceedings of the 2013 conference on Computer supported cooperative work companion - CSCW '13*. (2013), 315–320.

[50] Poikela, M. and Toch, E. 2017. Understanding the Valuation of Location Privacy: a Crowdsourcing-Based Approach. *Proceedings of the 50th Annual Hawaii International Conference on System Sciences*. (2017), 1985–1994.

[51] Quinn, K. and Epstein, D. 2018. #MyPrivacy: How Users Think About Social Media Privacy. *Proceedings of the 9th International Conference on Social Media and Society - SMSociety '18*. (2018), 360–364.

[52] Raynes-Goldie, K. 2010. Aliases, creeping, and wall cleaning: Understanding privacy in the age of Facebook. *First Monday*. 15, 1 (2010).

[53] Redmiles, E.M., Kross, S., Pradhan, A. and Mazurek, M.L. 2017. How Well Do My Results Generalize? Comparing Security and Privacy Survey Results from MTurk and Web Panels to the U.S. *University of Maryland Technical Reports of the Computer Science Department*. (2017).

[54] Spiekermann, S. and Cranor, L.F. 2009. Engineering privacy. *IEEE Transactions on Software Engineering*. 35, 1 (2009), 67–82.

[55] Stark, L. and Tierney, M. 2014. Lockbox : mobility , privacy and values in cloud storage. (2014), 1–13.

[56] Steinbart, P., Keith, M.J. and Babb, J.S. 2017. Measuring Privacy Concerns and the Right to Be Forgotten. (2017), 4967–4976.

[57] Strahan, R. and Gerbasi, K.C. 1972. Short, homogeneous versions of the Marlow-Crowne Social Desirability Scale. *Journal of Clinical Psychology*. 28, 2 (1972), 191–193.

[58] Straub, D., Boudreau, M.-C. and Gefen, D. 2004. Validation Guidelines for Is Positivist. *Communications of the Association for Information Systems*. 13, 24 (2004), 380–427.

[59] Stutzman, F. 2006. An evaluation of identity-sharing behavior in social network communities. *International Digital and Media Arts Journal*. 3, 1 (2006), 10–18.

[60] Suh, H., Shahriaree, N., Hekler, E.B. and Kientz, J.A. 2016. Developing and Validating the User Burden Scale. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*. (2016), 3988–3999.

[61] Vitak, J. 2012. The Impact of Context Collapse and Privacy on Social Network Site Disclosures. *Journal of Broadcasting and Electronic Media*. 56, 4 (2012), 451–470.

[62] WARNING: Google Buzz Has A Huge Privacy

Flaw: 2010. *https://www.businessinsider.com/warning-google-buzz-has-a-huge-privacy-flaw-2010-2*.

[63] Westin, A.F. 1967. *Privacy and Freedom*. New York: Atheneum.

[64] WhatsApp Status Update: Here's Why People Are Scared Of This Feature: 2017. *https://www.ndtv.com/offbeat/whatsapp-status-heres-why-people-are-scared-of-this-feature-1663139*.

[65] Wong, R.Y. and Mulligan, D.K. 2019. Bringing Design to the Privacy Table: Broadening "Design" in "Privacy by Design" Through the Lens of HCI. (2019).

[66] Wong, R.Y. and Mulligan, D.K. Bringing Design to the Privacy Table Broadening " Design " in " Privacy by Design " Through the Lens of HCI.

[67] Xu, H. 2007. The Effects of Self-Construal and Perceived Control on Privacy Concerns. *Twenty Eighth International Conference on Information Systems*. 6, 1 (2007), 1–14.

[68] Xu, H., Teo, H.-H., Tan, B.C.Y. and Agarwal, R. 2010. The Role of Push-Pull Technology in Privacy Calculus: The Case of Location-Based Services. *Journal of Management Information Systems*. 26, 3 (2010), 135–174.

[69] Young, A.L. and Quan-Haase, A. 2013. PRIVACY PROTECTION STRATEGIES ON FACEBOOK: The Internet privacy paradox revisited. *Information Communication and Society*. 16, 4 (2013), 479–500.

[70] Zhang, C. and Conrad, F.G. 2014. Speeding in Web Surveys : The tendency to answer very fast and its association with straightlining. 8, 2 (2014), 127–135.

10. Appendix

A Final Scale

	Ref.	In.	Institu.	Risk	Social
I think I have control over what personal information is shared by [X] with other companies.	[67]	Ct	0.97	0.07	-0.14
I believe I have control over how my personal information is used by [X].			0.79	0.01	0.03
I believe I have control over what personal information is collected by [X].			0.76	0.07	0.06
It is clear whether my personal information is shared with other companies.	[25]	Su	0.78	0.04	-0.06
I believe that [X] will prevent unauthorized people from accessing my personal information in their databases.			0.54	-0.15	0.21
I believe my personal information is accessible only to those authorized to have access.			0.71	-0.14	0.05
It is clear what information about me [X] keeps in their databases.	[4]	Tr	0.74	0.00	0.08
It is clear how long [X] retains my information.			0.77	0.16	-0.02
The purposes for which [X] asks for my information are clear.			0.77	-0.01	0.03
It is clear how [X] uses my personal information.	[25]		0.86	0.00	-0.02
I believe that if I would I ask, [X] will allow me to delete my personal information.	[56]	D	0.60	0.04	0.14
I think that it will be easy to delete my information from [X].			0.61	-0.03	0.18
I think it would be risky to give my personal information to [X].	[19]	R	-0.12	0.71	0.05
I think that there would be a high potential for privacy loss associated with giving my personal information to [X].			-0.04	0.67	0.03
My Personal information could be inappropriately used by [X].			-0.26	0.57	0.08
I think that providing [X] with my personal information would involve many unexpected problems.			0.08	0.82	0.00
I do not feel comfortable with the type of information I share using [X].			0.12	0.70	-0.13
Considering the information I provide to [X], and the people who might see it, I think it would be risky to give my personal information to [X].			0.12	0.80	-0.09
Considering the information I provide to [X], and the people who might see it, I think that there would be a high potential for privacy loss associated with giving my personal information to [X].			0.00	0.70	0.01
Considering the information I provide to [X], and the people who might see it, I think that providing [X] with my personal information would involve many unexpected problems.			0.03	0.79	0.09
I can understand whether people who I may know (friends, family, classmates, colleagues, acquaintances, etc.) have access to my personal information on [X].	[59]	Id	-0.12	0.10	0.72
It is clear who is the audience of my shared information on [X].			0.13	0.05	0.70
It looks easy to restrict un-intended people from viewing my personal information on [X].	[69]	Ps	0.09	-0.06	0.72
It looks easy to manage who can view my personal information on [X].			0.01	-0.07	0.73
I think [X] allows me to restrict the access to some of my personal information to some people.			-0.11	-0.06	0.75
I think I have control over what personal information is shared by [X] with other people.	[67]	Ct	0.22	0.06	0.63
It is clear what information about me others can see on [X].	[25]	Tr	0.13	0.05	0.70

Ct: Perceived information control, **Cf**: confidentiality, **Tr**: Importance of information transparency, **Su**: Secondary usage, **D**: Data deletion, **R**: Perceived privacy risk, **Is**: Information sensitivity, **Ps**: Protection strategies, **Id**: Identity sharing

B Gender and Age Distribution

Experiment	N	Gender distribution (%)			Age distribution (%)					
		Female	Male	Did not reveal	18-24	25-34	35-44	45-54	55-64	65+
Preliminary scale	241	80	158	3	34	138	43	16	9	1
Finalizing the scale	214	75	139		25	101	39	31	10	8
Using the scale	858	380	471	7	85	366	190	113	71	33

C Screening Task – First Two Experiments

Former studies in the field of decision making show that people, when making decisions and answering questions, are not always paying attention and are minimizing their effort as much as possible. A few studies show that over 50% of people don't carefully read questions. If you are reading this paragraph, in the first question please select the box marked 'other' and type 'evaluating information systems is fun' in the box below. Do not select anything else. In the second question, please select 'four'. Thank you for participating and taking the time to read through the questions carefully!

What was this study about? [Information systems evaluation, Making decisions about information systems, Evaluating information systems, Other]

It is common to evaluate information systems. [Strongly disagree (1), (2), (3), (4), Strongly agree (5)]

D Screening Task – Third Experiment

A few studies show that over 50% of people don't carefully read questions. If you are reading this paragraph, in the first question please select 'two'. In the second question, please select 'four'. Thank you for participating and taking the time to read through the questions carefully!

I usually take the time to evaluate information systems. [Strongly disagree (1), (2), (3), (4), Strongly agree (5)]

I think that evaluating information systems is important. [Strongly disagree (1), (2), (3), (4), Strongly agree (5)]

E Internal consistency and discriminant validity of constructs

				Factors correlations		
	Cronbach's α	AVE	SQRT(AVE)	Institutional	Social	Risk
Institutional	0.95	0.56	0.75		0.63	-0.23
Social	0.9	0.53	0.72			-0.29
Risk	0.9	0.53	0.73			

F Controlled Experiment: General Scenario Followed by One of the Two Cases

General scenario presentation: iFindRest

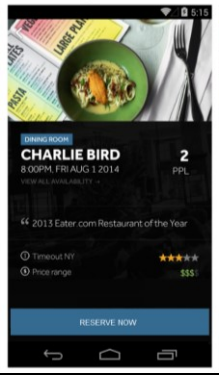
You are presented with a description of a future app, and we ask that you imagine yourself as a user in the specific scenario. Please read the description carefully and answer the following questions.

iFindRest

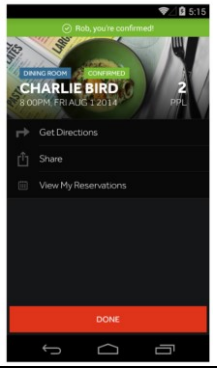
iFindRest is an app that helps with finding restaurants based on location and reserving a table in the desired restaurant.

The following screenshots demonstrate the app's user interface:

Reserving a table at the desired restaurant, at the desired time



Confirmation screen



Case 1: Privacy protective design

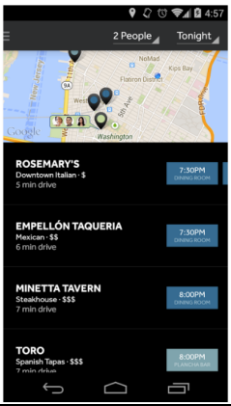
The scenario

Imagine that it is around 7:00 PM and you and your friend would like to go for a dinner at a nearby restaurant. You are using iFindRest to look for restaurants in your area, based on your current location. On the screen, you can see relevant restaurants. The restaurant that is marked in green indicates that other users, who are in your contact list, had made reservations to this restaurant at similar hours to yours. You cannot see who these users are since the default choice is not to share their identity publicly with their contact list, and they probably kept it as is.

Case 2: Privacy intrusive design

The scenario

Imagine that it is around 7:00 PM and you and your friend would like to go for a dinner at a nearby restaurant. You are using iFindRest to look for restaurants in your area, based on your current location. On the screen, you can see relevant restaurants. The restaurant that is marked in green indicates that other users, who are in your contact list, had made reservations to this restaurant at similar hours to yours. You can also see who these users are since the default choice is to share their identity publicly with their contact list, and they probably kept it as is.



General scenario presentation: iFit

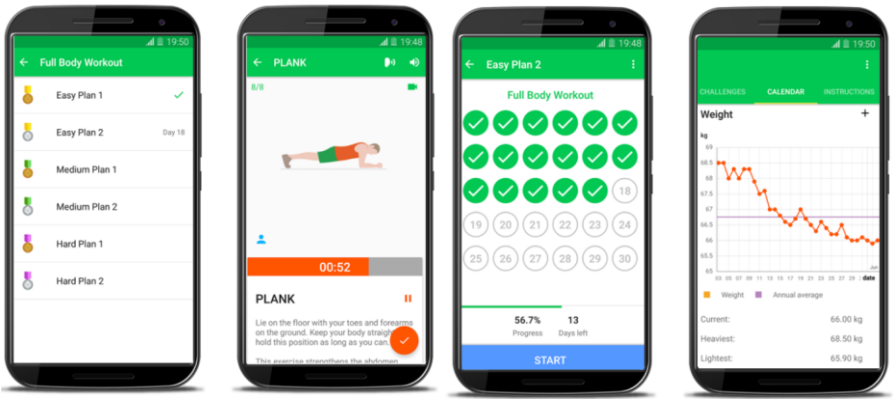
You are presented with a description of a future app, and we ask that you imagine yourself as a user in the specific scenario. Please read the description carefully and answer the following questions.

iFit

iFit is a fitness app which helps the users with doing exercises. The app provides a 30 days training programs for different parts of the body, at different difficulty levels.

The following screenshots demonstrate the app's user interface:

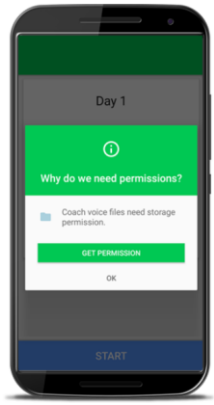
Choosing the desired difficulty level | Doing a specific exercise | Process recording | Weight recording



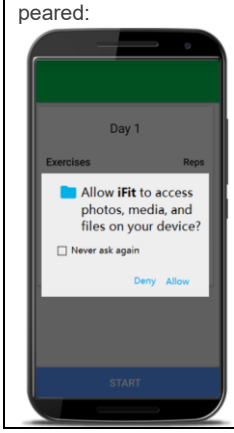
Case 1: Privacy protective design

The scenario

Imagine that this is the first time that you are using iFit. You chose "Easy Plan 1" which focuses on the abdominal muscles. You pressed "start" and the following message appeared:



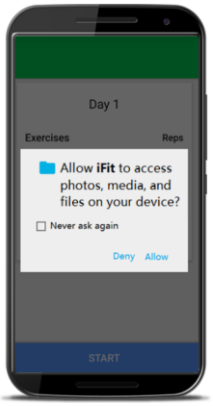
You clicked "GET PERMISSION" and the following message appeared:



Case 2: Privacy intrusive design

The scenario

Imagine that this is the first time that you are using iFit. You chose "Easy Plan 1" which focuses on the abdominal muscles. You pressed "start" and the following message appeared:



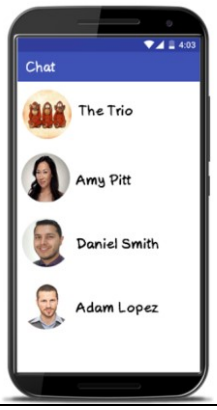
General scenario presentation: Message4All -Tale

You are presented with a description of a future app, and we ask that you imagine yourself as a user in the specific scenario. Please read the description carefully and answer the following questions.

Message4All
Message4All is a messenger app, similar to apps like WhatsApp, Snapchat, etc.

Users can chat with every contact on their phone. However, they are usually using the app for chatting with people with whom they have a close relationship, such as family, friends, and colleagues,
by sending text messages, photos, etc. The app is used for both one-on-one and groups chat conversations.

The following screenshot demonstrate the app's user interface:



Case 1: Privacy protective design

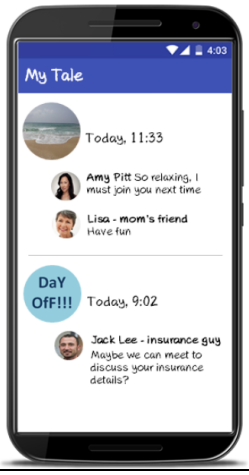
Tale is a feature in Message4All that allows the users to show content which can be seen by anyone who has the user's phone number and has Message4All installed. The content will be available for 24 hours only and will be automatically dismissed afterward.

The scenario
Imagine that you decided to try the Tale feature. You took a day off and were about to share a video showing the beach you went to. After pressing the "Share" button, the following message appeared on the screen:

Case 2: Privacy intrusive design

Tale is a feature in Message4All that allows the users to show content which can be seen by anyone by default, which has the user's phone number and has Message4All installed. The content will be available for 24 hours only and will be automatically dismissed afterward.

The scenario
Imagine that you decided to try the Tale feature. You took a day off and shared two tales. The first one was a text tale and the second contained a video of the beach you went to. During the day, few people commented on your tales, with some of them you rarely speak. You can see your tales and their comments as demonstrated in the following screenshot:



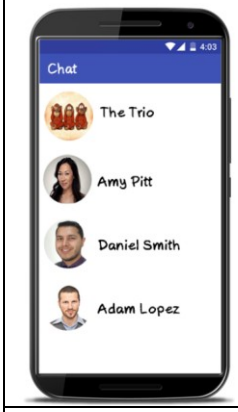
General scenario presentation: Message4All - Groups

You are presented with a description of a future app, and we ask that you imagine yourself as a user in the specific scenario. Please read the description carefully and answer the following questions.

Message4All
Message4All is a messenger app, similar to apps like WhatsApp, Snapchat, etc.

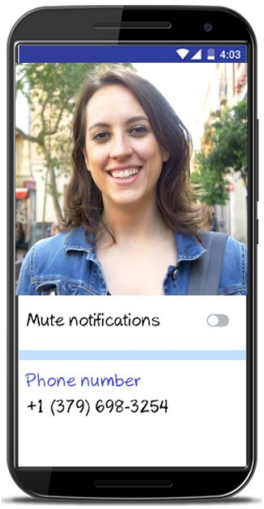
Users can chat with every contact on their phone. However, they are usually using the app for chatting with people with whom they have a close relationship, such as family, friends, and colleagues, by sending text messages, photos, etc. The app is used for both one-on-one and groups chat conversations.

The following screenshot demonstrate the app's user interface:



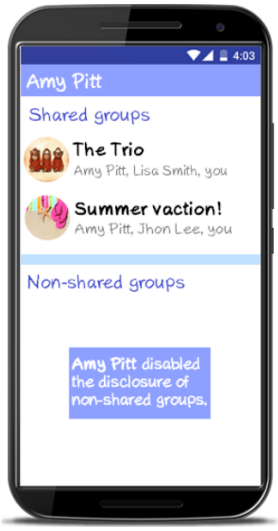
Contact person details
Within Message4All contact list, a user can get further information about specific contact person and set settings. For example, the user can review the groups that the contact person is part of, both groups that they have in common and also those they do not share.

The following screenshot demonstrates the app's contact person interface:



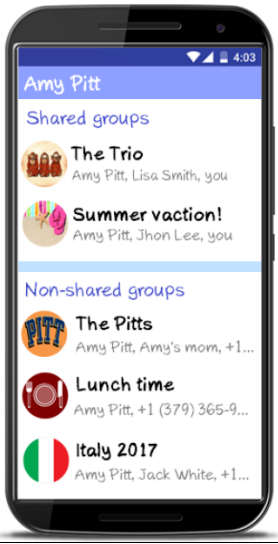
Case 1: Privacy protective design

The scenario
Imagine that you have a friend named Amy Pitt and you often chat with her using Message4All. You wanted to look for a group that you remembered that you are both members of. Therefore, you looked at her details on the app. Here is a screenshot that provides information about Amy's groups.



Case 2: Privacy intrusive design

The scenario
Imagine that you have a friend named Amy Pitt and you often chat with her using Message4All. You wanted to look for a group that you remembered that you are both members of. Therefore, you looked at her details on the app. Here is a screenshot that provides information about Amy's groups.



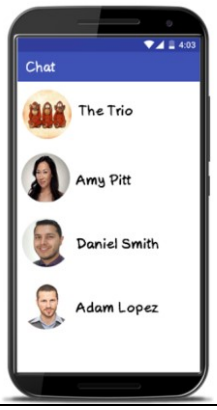
General scenario presentation: Message4All - Ad

You are presented with a description of a future app, and we ask that you imagine yourself as a user in the specific scenario. Please read the description carefully and answer the following questions.

Message4All
Message4All is a messenger app, similar to apps like WhatsApp, Snapchat, etc.

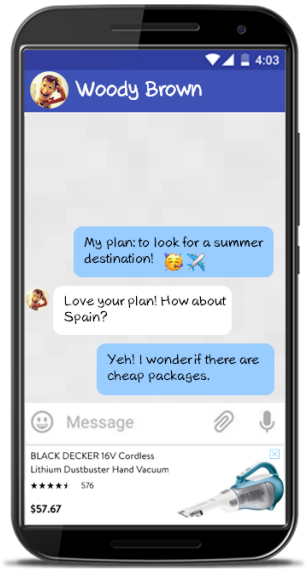
Users can chat with every contact on their phone. However, they are usually using the app for chatting with people with whom they have a close relationship, such as family, friends, and colleagues, by sending text messages, photos, etc. The app is used for both one-on-one and groups chat conversations.

The following screenshot demonstrate the app's user interface:



Case 1: Privacy protective design

The scenario
Imagine that you are using Message4All to chat with your friend Woody. Here is the screenshot of your chat:



Case 2: Privacy intrusive design

The scenario
Imagine that you are using Message4All to chat with your friend Woody. Here is the screenshot of your chat: