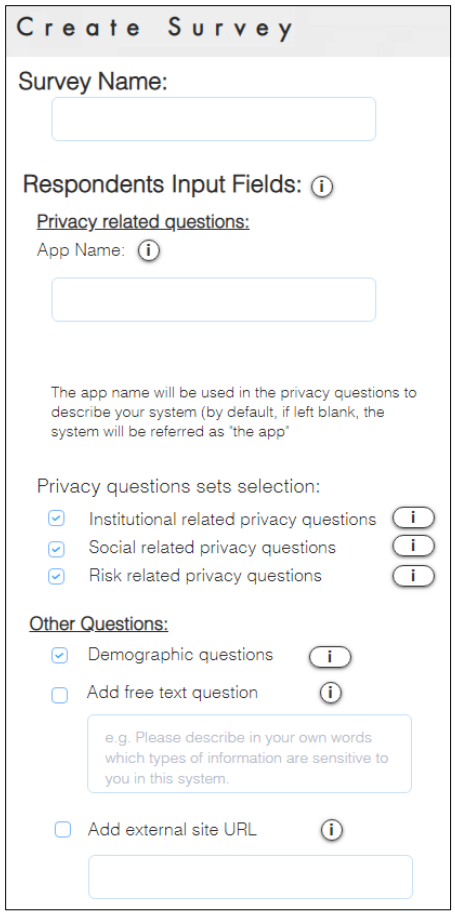

Figure 1: A partial screenshot of our A/P(rivacy) Testing platform, demonstrating part of the researchers' interface, in which they start setting the desired experiment.

A/P(rivacy) Testing: Assessing Applications for Social and Institutional Privacy

Oshrat Ayalon
Department of industrial engineering
Tel Aviv University
Tel Aviv, Israel
oshratra@tauex.tau.ac.il

Eran Toch
Department of industrial engineering
Tel Aviv University
Tel Aviv, Israel
erant@tauex.tau.ac.il

ABSTRACT

The way information systems are designed has a crucial effect on users' privacy, but users are rarely involved in Privacy-by-Design processes. To bridge this gap, we investigate how User-Centered Design (UCD) methods can be used to improve the privacy of systems' designs. We present the process of developing A/P(rivacy) Testing, a platform that allows designers to compare several privacy designs alternatives, eliciting end-users' privacy perceptions of a tested system or a feature (Figure 1). We describe three online experiments, with 959 participants, in which we created and validated the reliability of a scale for Users' Perceived Systems' Privacy (UPSP), and used it to compare between privacy designs alternatives by using scenarios and different variants. We show that A/B testing is applicable for privacy purposes and that our scale is differentiating between designs that perceived as legitimate and designs that may violate users' expectations.

CCS CONCEPTS

• **Security and privacy** → **Human and societal aspects of security and privacy** → Usability in security and privacy

CHI'19 Extended Abstracts, May 4-9, 2019, Glasgow, Scotland, UK.

ACM ISBN 978-1-4503-5971-9/19/05.
DOI: https://doi.org/10.1145/3290607.3312972

KEYWORDS

Privacy; user-centered design; privacy-by-design; A/B testing; controlled experiments

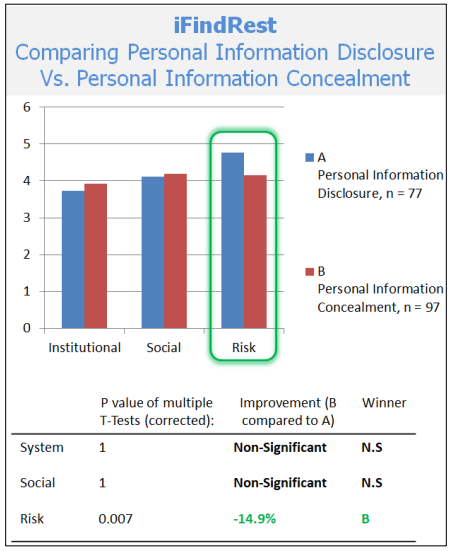

Figure 2: A visualization that can be generated by the researcher using the A/P testing output, comparing A and B, two variants of a design.

INTRODUCTION

System designs that do not meet the users' privacy expectations can startle users and lead them to abandon the system altogether [8]. Privacy-by-design (PbD) initiatives propose a design and development framework that aids in the production of privacy-respectful systems [5]. While PbD is part of official guidelines by the FTC and by the recent European General Data Protection Regulation (GDPR) [9], it is also criticized of being too focused on compliance to privacy regulation, rather than on answering users' privacy expectations [12].

When designing for privacy, developers usually consult with other developers or with the Chief Privacy Offices (CPOs) [3, 10]. However, leaning only on internal feedback before launching a new system or a feature can end up in systems that mismatch users' privacy expectations [4]. This is particularly important as end-users' privacy expectations are not only based on the way their data is handled between them and the system (an aspect known as institutional privacy [16, 17]). Rather, expectations also cover social privacy: how systems allow managing relationships between end-users, and the complexity that sharing and hiding information plays in these relationships [13, 16, 17].

The motivation of the study is helping developers with developing information systems that are more aligned with the users' privacy expectations. However, first, there is a need to know what these expectations are. Therefore, the study goal is developing tools for determining the users' perceptions of a new or existing information systems' privacy. To reach this goal, and based on current gaps in literature, we developed and evaluated Users' Perceived Systems' Privacy (UPSP) scale. Our developed scale adds a social aspect that highlights the information flow between people using the system and compared it to scales that reflected the flow of information between the system and the individual user. To demonstrate how the scale can be used, we developed A/P(rivacy) Testing, a tool for determining the users' perceptions of the information systems' privacy. In Figure 2 we present a visualization of the results that can be generated by the researcher based on the A/P testing output.

SCALE DEVELOPMENT AND RELIABILITY VALIDATION

The first stage of the study included two experiments in which we developed the scale and validated its reliability. In the first experiment we tested our initial questionnaire which was consisted of 47 initial items. The items were divided to two types of questions: Twenty-seven questions were institutional-related and 20 questions were social-related. The initial items are both based on previous studies [1, 6, 7, 11, 18–21] and also were developed by us. At the end of the survey the participants answered two screening questions, and participants considered as qualified if they answered both questions correctly. We recruited 300 participants (final sample size: 241) via Amazon Mechanical Turk (AMT) and performed principle component analyses (PCA) and exploratory factor analyses (EFA) to analyze the results. The EFA results pointed to three distinct

Table 1: Several examples of the final questions as they appear in the Users' Perceived Systems' Privacy (UPSP) scale. *sub-scale topic: I = institutional, S = social, R = risk

*	Statement
I	I think I have control over what personal information is shared by [X] with other companies.
I	I believe I have control over how my personal information is used by [X].
I	I believe I have control over what personal information is collected by [X].
S	It looks easy to share content on [X] with specific people.
S	I can understand whether people who I may know (friends, family, classmates, colleagues, acquaintances, etc.) have access to my personal information on [X].
S	It is clear who is the audience of my shared information on [X].
R	I do not feel comfortable with the type of information [X] requests from me.
R	I do not feel comfortable with the type of information I share using [X].
R	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].

constructs. We removed questions with loading value lower than 0.5, which resulted in a questionnaire that is consisted of 33 questions, divided to three sub-scales: institutional-related, social-related and risk-related. The risk questions were consisted of questions related to both institutional and social aspects (our primary division). In the second experiment we recruited 300 participants (final sample size: 218) to finalize the scale. We used Confirmatory Factor Analysis (CFA) to validate the scale's reliability. Our goodness-of-fit analyses showed our data supported the model of 33 questions distributed in the previously found three sub-scales: $\chi 2/df = 1.87$, RMSEA = 0.063, SRMR = 0.059, CFI = 0.92, TLI = 0.91. In addition, based on the second experiment we also changed the wordings of some of the questions. See

Table **1** for final questions examples.

TESTING DESIGN VARIANTS

In the second stage of study we explored how we can use the scale to compare several privacy design variations.

Method

We designed a between-subject user study, using an online experiment that included a scenario presentation followed by the UPSP scale. We created three scenarios and per each scenario we created two cases, differing in their privacy design: privacy intrusive design versus privacy respective. Altogether, we had three background scenarios and six cases. We recruited 600 participants (final sample size: 500), and they were randomly assigned to one of the six scenario-case combinations only.

We designed the general scenarios and the cases based on our previous study [2]. We 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 scale questions. The statements were presented as three sub-scales: institutional, social, and risk. The sub-scales were ordered accordingly, and the statements within each subscale were randomly ordered.

General scenario: Message4All, Tale feature. A user can share content with all the app's users who have his/her phone number, for a limited time.

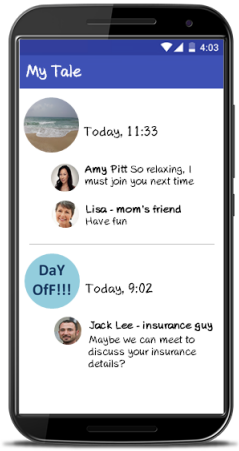

Figure 3: Intrusive privacy design. The user automatically shares the information with the entire contact list.

Figure 4: Respective privacy design. The user first chooses with whom to share the information.

As we were interested in testing the social aspect of information systems, two apps and their described scenarios had a prominent social aspect, and one app and its described scenario was not focused on a social issue. The applications' names were invented, but we have based the applications' functionalities on existing applications. The three applications and features used were: 1) iFit, a fitness app which helps the users with doing exercises; 2) iFindRest, which helps the users with finding restaurants based on their location and reserving a table; 3) 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. Figure 3 and Figure 4 demonstrates the screenshots of the two designs of the Tale feature in Message4All app, in which Figure 3 is privacy intrusively designed and Figure 4 is privacy respectively designed.

Results

To analyze our results we compared 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 between the two cases per each sub-scale. Thus, we performed T-tests and used Bonferroni correction for multiple comparisons.

Our results showed that within part of the different scenarios, some of the differences between the cases were significant and are summarized in Table 2. In the scenario that referred to Message4All app, and included a specific feature Tale, we found significant differences between the cases for all the subscales ($p < 0.01$). The privacy respective design was perceived as so for all categories: it was more privacy respective from the system and social aspects and perceived as less risky. In addition, the social aspect had the greatest size effect (Cohen's d = 0.92). In the iFindRest scenario, the intrusive design was perceived as riskier compared to the respective design ($p = 0.01$), and we did not find significant differences in the other categories. Lastly, we did not find significant differences between the cases in the iFit scenario for any of the categories. Figure 5 summarizes the mean sub-scales scores of each scenario, comparing between the two cases.

DISCUSSION

We were motivated by the Privacy by Design (PbD) approach and encouraged by inclusion of PbD in the May 2018 European GDPR. However, PbD can be criticized in a similar way that mainstream system design was criticized by the User Centered Design approach [14]. 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 [15]. Our results point to particular contexts in which system design can be considered as inappropriate. Specifically, issues of social privacy bother users the most and they expect systems to provide them with defaults and controls that would follow their social privacy expectations. We borrowed the A/B testing methodology and our findings show that this methodology can be applied to the field of privacy.

Table 2: Comparing the cases per each scenario, exploring in which subscales there are significant differences in the mean score. * 0.01, ** <0.01

Scen.	Sub - scale	Res.	Int.	Cohen's d
iFind	instit.	3.93	3.74	0.13
Rest	social	4.19	4.11	0.05
	***risk**	**4.15**	**4.77**	**0.47**
iFit	instit.	3.42	3.12	0.21
	social	3.26	3.06	0.13
	risk	4.82	4.93	0.08
Msg4All	****instit.**	**4.48**	**3.74**	**0.56**
	****social**	**5.32**	**4.19**	**0.92**
	****risk**	**4.10**	**4.81**	**0.49**

In addition, in order to conduct the experiment we developed a scale to measure users' perceived privacy of a tested information system. The novelty of our scale is its multi-facets, covering both social and institutional privacy, and its approach, aimed to evaluate systems' privacy as it perceived by the users, rather than individuals' general privacy concerns. Through our work, we suggested a framework for evaluating users' perceived privacy, as we described in section 3.1. The framework is necessary to demonstrate privacy issues in a simple and a concise way, 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); *versions level*: 4) Case description. 5) Answering the UPSP scale.

Based on our study results we developed the assistive tool A/P(rivacy) testing. The tool allows researchers and decision-makers in the industry to evaluate information systems based on our suggested framework. The tool has two different interfaces: 1) for the researcher and 2) the respondents. The researcher interface includes fields to fill in according to our suggested framework, creating several versions of a tested system. We provided the scale questions, divided into the three distinct subscales as were found (institutional, social, risk). The researchers may choose which subscale to include and also which questions, in case they want to choose several questions only, with the highest loading values, for example. In addition to providing the information about the tested system and the versions, the researchers may add other questions and to use an external service to add other questions. The respondent's interface is similar to existing surveys platforms, in which the respondent is presented with the general information, specific case, and the scale questions, as were chosen by the researcher. The advantage of using the tool is in its simplicity and the embedded framework, allowing the creation of several versions to be automatically randomly assigned to the respondents. Figure 1 demonstrates a screenshot of the researchers' interface.

FUTURE WORK

Our developed tool opens varying options for future studies. For example, the GDPR strongly refers to informed consent, requiring data controllers to make sure that the data subjects understand how their data is being processed. However, many times users choose "agree," but they do not know what their choice is referring to. A future study may explore different ways to make the consent form more visualized and investigate its effect over users' perceived privacy.

ACKNOWLEDGMENTS

This work was supported by the Israeli Ministry of Science and Technology (Shulamit Aloni grant, number 314575 and international conference travel grant, number 315768) and by the ICRC – Blavatnik Interdisciplinary Cyber Research Center, grant number 590713. We would also like to

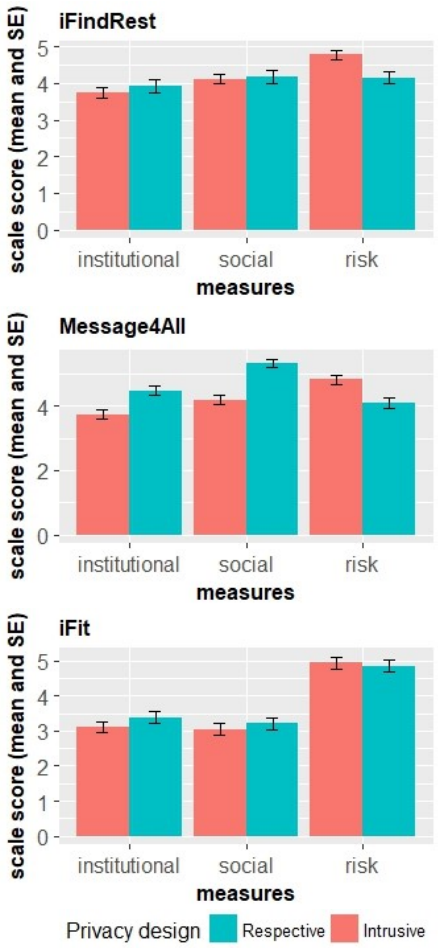

Figure 5: Per each scenario (iFindRest, Message4All, iFit), we compared the two privacy designs, respective vs. intrusive, per each subscale measure

thank Luiza Jarovsky for helping us with finalizing the scale and Shany Peter for developing A/P Testing.

REFERENCES

[1] 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.

[2] Ayalon, O. and Toch, E. 2018. Crowdsourcing Privacy Design Critique: An Empirical Evaluation of Framing Effects. *Hawaii International Conference on System Sciences 2018*. (2018), 4752–4761.

[3] Balebako, R., Cranor, L. and Mellon, C. 2014. Improving App Privacy: Nudging App Developers to Protect User Privacy. (2014), 55–58.

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

[5] Cavoukian, A. 2009. Privacy by design: The 7 foundational principles. Information and Privacy Commissioner of Ontario, Canada.

[6] 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.

[7] 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.

[8] 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.

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

[10] 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.

[11] Hong, W. 2013. R ESEARCH N OTE I NTERNET P RIVACY C ONCERNS: A N I NTEGRATED Example Items from Existing IPC Instruments. 37, 1 (2013), 1–3.

[12] 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 and Technology*. 28, 2 (2014), 159–171.

[13] 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.

[14] 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.

[15] Nissenbaum, H. 2009. *Privacy in context: Technology, policy, and the integrity of social life*. Stanford University Press.

[16] 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.

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

[18] Steinbart, P., Keith, M.J. and Babb, J.S. Measuring Privacy Concerns and the Right to Be Forgotten.

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

[20] 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.

[21] 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.